National Library of Canada

Acquisitions and Bibliographic Services Branch

395 Wellington Street
Ottawa, Ontario
K1A 0N4

Bibliothèque nationale du Canada

Direction des acquisitions et des services bibliographiques

395, rue Wellington
Ottawa (Ontario)
K1A 0N4

Your file Votre référence

Our file Notre référence

NOTICE

The quality of this microform is heavily dependent upon the quality of the original thesis submitted for microfilming. Every effort has been made to ensure the highest quality of reproduction possible.

If pages are missing, contact the university which granted the degree.

Some pages may have indistinct print especially if the original pages were typed with a poor typewriter ribbon or if the university sent us an inferior photocopy.

Reproduction in full or in part of this microform is governed by the Canadian Copyright Act, R.S.C. 1970, c. C-30, and subsequent amendments.

AVIS

La qualité de cette microforme dépend grandement de la qualité de la thèse soumise au microfilmage. Nous avons tout fait pour assurer une qualité supérieure de reproduction.

S'il manque des pages, veuillez communiquer avec l'université qui a conféré le grade.

La qualité d'impression de certaines pages peut laisser à désirer, surtout si les pages originales ont été dactylographiées à l'aide d'un ruban usé ou si l'université nous a fait parvenir une photocopie de qualité inférieure.

La reproduction, même partielle, de cette microforme est soumise à la Loi canadienne sur le droit d'auteur, SRC 1970, c. C-30, et ses amendements subséquents.

Canada

UNIVERSITY OF ALBERTA

THE MOLECULAR STRUCTURES OF TWO CYTOTOXIC RIBONUCLEASES OF VERTEBRATES, ONCONASE AND EOSINOPHIL-DERIVED NEUROTOXIN

by

STEVEN C. MOSIMANN

A THESIS

SUBMITTED TO THE FACULTY OF GRADUATE STUDIES AND RESEARCH

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE

DEGREE OF DOCTOR OF PHILOSOPHY

DEPARTMENT OF BIOCHEMISTRY

EDMONTON, ALBERTA

FALL 1995

National Library
of Canada

Acquisitions and
Bibliographic Services Branch

395 Wellington Street
Ottawa, Ontario
K1A 0N4

Bibliothèque nationale
du Canada

Direction des acquisitions et
des services bibliographiques

395, rue Wellington
Ottawa (Ontario)
K1A 0N4

Your file Votre référence

Our file Notre référence

THE AUTHOR HAS GRANTED AN IRREVOCABLE NON-EXCLUSIVE LICENCE ALLOWING THE NATIONAL LIBRARY OF CANADA TO REPRODUCE, LOAN, DISTRIBUTE OR SELL COPIES OF HIS/HER THESIS BY ANY MEANS AND IN ANY FORM OR FORMAT, MAKING THIS THESIS AVAILABLE TO INTERESTED PERSONS.

L'AUTEUR A ACCORDE UNE LICENCE IRREVOCABLE ET NON EXCLUSIVE PERMETTANT A LA BIBLIOTHEQUE NATIONALE DU CANADA DE REPRODUIRE, PRETER, DISTRIBUER OU VENDRE DES COPIES DE SA THESE DE QUELQUE MANIERE ET SOUS QUELQUE FORME QUE CE SOIT POUR METTRE DES EXEMPLAIRES DE CETTE THESE A LA DISPOSITION DES PERSONNE INTERESSEES.

THE AUTHOR RETAINS OWNERSHIP OF THE COPYRIGHT IN HIS/HER THESIS. NEITHER THE THESIS NOR SUBSTANTIAL EXTRACTS FROM IT MAY BE PRINTED OR OTHERWISE REPRODUCED WITHOUT HIS/HER PERMISSION.

L'AUTEUR CONSERVE LA PROPRIETE DU DROIT D'AUTEUR QUI PROTEGE SA THESE. NI LA THESE NI DES EXTRAITS SUBSTANTIELS DE CELLE-CI NE DOIVENT ETRE IMPRIMES OU AUTREMENT REPRODUITS SANS SON AUTORISATION.

ISBN 0-612-06261-9

Canada

UNIVERSITY OF ALBERTA

RELEASE FORM

NAME OF AUTHOR: Steven Craig Mosimann

TITLE OF THESIS: The Molecular Structures of two Cytotoxic Ribonucleases of Vertebrates, Onconase and Eosinophil-Derived Neurotoxin

DEGREE: Doctor of Philosophy

YEAR THIS DEGREE GRANTED: 1995

Permission is hereby granted to the University of Alberta Library to reproduce single copies of this thesis and to lend or sell such copies for private, scholarly or scientific research purposes only.

The author reserves all other publication and other rights in association with the copyright in the thesis, and except as hereinbefore provided neither the thesis nor any substantial portion thereof may be printed or otherwise reproduced in any material form whatever without the author's prior written permission.

S Mosimann

12130 - 140 Ave, Edmonton, AB. T5X 5K4

October 6, 1995

UNIVERSITY OF ALBERTA

FACULTY OF GRADUATE STUDIES AND RESEARCH

The undersigned certify that they have read, and recommend to the Faculty of Graduate Studies and Research for acceptance, a thesis entitled "The Molecular Structures of two Cytotoxic Ribonucleases of Vertebrates, Onconase and Eosinophil-Derived Neurotoxin" submitted by Steven C. Mosimann in partial fulfillment of the requirements for the degree of Doctor of Philosophy.

Michael James
Dr. Michael N.G. James

Cyril M. Kay
Dr. Cyril Kay

Joel Weiner
Dr. Joel Weiner

John Vederas
Dr. John Vederas

S R Sprang
Dr. Stephen Sprang

Date: October 6, 1995

ABSTRACT

Onconase™ (P-30 Protein; ONC) is a cytotoxic, pyrimidine-specific ribonuclease of *Rana pipiens* with therapeutic properties *in vivo*. Orthorhombic single crystals of purified ONC (space group $P2_12_12_1$) are obtained from an $(NH_4)_2SO_4$ solution at pH 4.5 and grow as rosettes: a = 40.53 (8), b = 69.64 (9), c = 32.52 (8) Å. The crystal structure of ONC has been determined by multiple-isomorphous replacement, using three heavy atom derivatives and refined at 1.7 Å resolution. The comparative molecular model of ONC based upon the known structure of ribonuclease A (RNase A) and the subsequent crystal structure demonstrate ONC can adopt an RNase A-like polypeptide fold. However, there are many large conformational differences in the large loop structures of ONC, the comparative molecular model and RNase A. Unlike RNase A, the N-terminal residue (pyroglutamyl) of ONC hydrogen bonds to Lys9N$^{\zeta}$ at the active site of the enzyme. Recombinant Onconase (rONC) having an N-terminal sequence $^{f}Met_{-1}Glu_1Asp_2\cdots$ has been expressed, purified and shown to lack significant cytotoxic and ribonucleolytic activity. The isomorphous rONC structure has been determined and refined at 2.2 Å resolution to examine the structural basis of this inactivity. Large conformational differences in the ONC and rONC structures are confined to the N-terminal residues and only slightly perturb the conformation of the active site residues. Molecular models of ONC and rONC-d(UpG) complexes have been constructed and attribute the lack of ribonucleolytic activity of rONC to differences in the shape and electrostatic nature of the ribose binding site (R_2). The modeling implicates Lys33 in the uridine base specificity of Onconase and suggests Ser54, Asn56 and Glu91 are responsible for binding the purine base.

In parallel work, the X-ray crystallographic structure of the related recombinant eosinophil-derived neurotoxin (rEDN) has been determined and refined at 1.83Å resolution. This structure reveals the conformation of several novel insertions unique to

EDN and the presence of a second sulfate anion binding site. The additional anion binding site may correspond to the P_{-1} phosphate binding subsite for oligoribonucleotide substrates.

Acknowledgments

I thank my collaborators for providing samples of purified protein and making this work possible. Onconase was isolated, purified and sequenced by W. Ardelt of the Alfacell Corporation. Recombinant Onconase and recombinant eosinophil-derived neurotoxin were expressed, isolated and purified by Y. Wu and D. Newton, respectively, in the laboratory of R.Youle. Within our laboratory, I thank K. Johns for the initial crystallization of Onconase and N. Strynadka for assistance with an early data collection. This work has been supported by the Medical Research Council of Canada (M.James), a University of Alberta, Faculty of Medicine Graduate Studentship (SM) and an Alberta Heritage Fund for Medical Research Graduate Studentship (SM).

Throughout my graduate studentship the members of the 'James gang' have been warmhearted, understanding and talented. In particular, I am indebted to Marie, Natalie and Anita for sharing their views about science, crystallography and life in general. More recently, I have had valuable discussions with Ernst and Masao. I also thank Maia for her assistance in the wet lab and Ron for his aid with computational problems. To my fellow graduate students Stan, Kui, Kathy, Nina and Amir, thanks for the memories.

Finally, I would like to thank M. Wylie for her help in the preparation of this thesis and my supervisor Michael N.G. James. Mike takes a great deal of pride in his work and expects a great deal from members of his lab. In many ways, the quality of this work can be attributed to Mike's ability to spot the small inconsistencies that creep into manuscripts and figures. I am particularly grateful for his understanding and support through good times and bad in the lab and in life.

TABLE OF CONTENTS

Page

CHAPTER 1: INTRODUCTION 15

Pyrimidine Specific RNase Superfamily of Vertebrates 16

Cytotoxic Ribonucleases 18

Non-Cytotoxic Ribonucleases 23

Ribonucleolytic Mechanism 25

Enzymatic Activity 26

References 32

CHAPTER 2: COMPARATIVE MOLECULAR MODELING AND CRYSTALLIZATION OF P-30 PROTEIN: A NOVEL ANTI-TUMOR PROTEIN OF *Rana pipiens* OOCYTES AND EARLY EMBRYOS 38

INTRODUCTION 38

RESULTS 39

Alignment of P-30 and RNase A Sequence 39

Model Building and Energy Minimization 41

RNase A and Energy Minimized RNase A 44

P-30 Comparative Molecular Model 46

Crystallization 54

REFERENCES 55

Page

CHAPTER 3: REFINED 1.7 Å X-RAY CRYSTALLOGRAPHIC STRUCTURE OF P-30 PROTEIN, AN AMPHIBIAN RIBONUCLEASE WITH ANTI-TUMOR ACTIVITY 58

INTRODUCTION 58

METHODS 59

Structure Determination 59

Structure Refinement 63

RESULTS 64

Quality of the Refined Structure 64

Molecular Conformation of P-30 Protein 72

Intermolecular Contacts 77

Active Site of P-30 Protein 81

DISCUSSION 85

REFERENCES 90

CHAPTER 4: REFINED 2.2 Å X-RAY CRYSTALLOGRAPHIC STRUCTURE OF RECOMBINANT ONCONASE AND COMPARISON WITH RNASE A AND ONCONASE 94

INTRODUCTION 94

RESULTS 95

Quality of the Refined Structure 95

Molecular Conformation of rONC 98

Superposition of rONC And ONC with RNase A 104

DISCUSSION 111

MATERIALS AND METHODS 115

Protein Purification and Crystallization 115

Data Collection 115

Refinement 116

REFERENCES 120

Page

CHAPTER 5: X-RAY CRYSTALLOGRAPHIC STRUCTURE OF RECOMBINANT EOSINOPHIL-DERIVED NEUROTOXIN AT 1.83 Å RESOLUTION 123

INTRODUCTION 123

METHODS 124

RESULTS 128

Quality of the Refined Structure 128

Molecular Conformation of rEDN 132

Sulfate Anion Binding Sites 141

Intermolecular Packing in rEDN 144

Superposition with RNase A and Onconase 147

DISCUSSION 150

REFERENCES 157

CHAPTER 6: DISCUSSION 161

B_1 Subsite of Pyrimidine Specific RNases 162

B_2 subsite of Pyrimidine Specific RNases 163

rONC 163

P_{-1} Subsite of rEDN 164

Comparative Molecular Modeling 164

References 166

LIST OF TABLES

		Page
Table 1.1	Pyrimidine-Specific RNase Superfamily of Vertebrates	17
Table 1.2	Amino Acid Sequence Alignment of Selected Members of the RNase Superfamily	19
Table 1.3	Amino Acid Sequence Identity between Members of the RNase Superfamily	20
Table 1.4	Enzymatic Activity and Specificity of Members of the RNase Superfamily	30
Table 2.1	Complete Amino Acid Sequence Alignment of Pancreatic RNase A and P-30 Protein	40
Table 2.2	Rms Deviation in Coordinates For Minimized RNase A and the P-30 Protein Model	47
Table 3.1	Summary of Data Collection Statistics for P-30 Native and Derivative Data Sets	61
Table 3.2	Multiple Isomorphous Replacement Phasing Statistics for P-30 at 2.7Å Resolution	62
Table 3.3	Stereochemical Statistics for the Final Cycle of Restrained Least-Square Refinement	65
Table 3.4	Reverse Turns in P-30	69
Table 3.5	Intermolecular Close Contacts and Hydrogen Bonds for the Refined P-30 Model	78
Table 4.1	Stereochemical Statistics for the Final Cycle of Least-Square Refinement of rONC	96
Table 4.2	Conserved Hydrogen Bonds Involving the Sulfate Anion in the Superposed ONC and rONC Structures	103
Table 4.3	Summary of Data Collection Statistics for the Onconase(Met-1,Glu1) Native Data Set	117
Table 5.1	Summary of Data Collection Statistics for the rEDN Native Data Set	126

		Page
Table 5.2	Stereochemical Statistics for the Final Cycle of Restrained Least-Squares Refinement	129
Table 5.3	Secondary Structures of rEDN	138
Table 5.4	β-Turn Structures of rEDN	139
Table 5.5	The Sulfate Anion Contacts Distances (Å) in rEDN	143
Table 5.6	The Intermolecular Contacts in EDN	146

LIST OF FIGURES

		Page
Figure 1.1	The Catalytic Mechanism of Pyrimidine Specific RNases	27
Figure 1.2	The Nomenclature of the Base, Ribose and Phosphate Moieties of RNA Substrates and Analogs	29
Figure 2.1a	An Example of Sequence Alignment Revision based upon Inspection of RNase A is Shown (Automated Sequence Alignment)	42
Figure 2.1b	An Example of Sequence Alignment Revision Based Upon Inspection of RNase A is Shown (Revised Structural Alignment)	42
Figure 2.2	C^{α} Representation of RNase A Showing the Deletions Relative to the P-30 Protein Alignment In Table 2.1	43
Figure 2.3	An Extreme Example of Unacceptable Close Contacts and its Relief through Graphical Intervention	45
Figure 2.4	Superposition of RNase A Crystallographic Structure before and after Energy Minimization	48
Figure 2.5	Superposition of Minimized RNase A and the P-30 Protein Model	49
Figure 2.6	The Tentative P-30 Protein Model	53
Figure 3.1	The $2\|F_o\|-\|F_c\|$ Electron Density (1σ Contour) Associated with Arg40-Glu42 and Thr25'-Leu27'	67
Figure 3.2	Ramachandran Plot of the Completed P-30 Model	68
Figure 3.3	The $2\|F_o\|-\|F_c\|$ Electron Density Belonging To Gln93	70
Figure 3.4	A Plot of the Average Thermal Parameter (B-Factor) against the Amino Acid Sequence Number of P-30	71
Figure 3.5a	A Molscript Ribbons Diagram of the Molecular Conformation of the P-30 Protein	73
Figure 3.5b	An All Atom Representation of the P-30 Molecule	73
Figure 3.5c	The Main-Chain Tracing of the P-30 Molecule	73

Page

Figure 3.6	The Intramolecular Main-Chain Hydrogen Bonding Network of P-30	74
Figure 3.7	The Intermolecular Packing Diagram of Orthorhombic P-30	79
Figure 3.8	The Intermolecular Contact at the Active Site of P-30	80
Figure 3.9	The Electron Density Associated with the Active Site	82
Figure 3.10	An All Atom Representation of the Active Site of P-30 is Shown with the Detailed Hydrogen Bonding Network	83
Figure 3.11	The C^{α} Superposition of the Comparative Molecular Model and X-Ray Crystallographic Structure of P-30	86
Figure 3.12	The Predicted and Real Conformations of the P-30 Loop Containing Residues Lys49-Asn54	88
Figure 4.1	The $3\|F_o\|-2\|F_c\|$ Electron Density Associated with the N-Terminal α-Helix of rONC	97
Figure 4.2	The $\|F_o\|-\|F_c\|$ Electron Density Associated with the Sulfate Anion Site After Refinement Leaving out the Atoms of the Anion ..	99
Figure 4.3	A Representation of the Main-Chain Atoms of the Least-Squares Superposed Structures of ONC and rONC	100
Figure 4.4	An All Atom Representation of the Active Sites of the Least-Squares Superposed Structures of ONC and rONC	102
Figure 4.5	The Least-Squares Superposition of Lobe 1 of rONC and the Equivalent Atoms in RNase A	105
Figure 4.6	The Superposed ONC and Sulfate Containing RNase A Active Sites Using the Superposition from Figure 4.5	107
Figure 4.7a	The Active Site of the ONC-d(UpG) Model Based on RNase A-d(CpA)	109
Figure 4.7b	The rONC-d(UpG) Model	110
Figure 5.1a	The Omit Electron Density Associated with the Sulfate Anions	130

		Page
Figure 5.1b	Electron Density Associated with Active Site Residues of rEDN	131
Figure 5.2	The Average Main-Chain and Side-Chain B-Factor versus Sequence Number Plots for the rEDN Structure	133
Figure 5.3	A Ramachandran Plot of the ϕ, ψ Dihedral Angles of rEDN	134
Figure 5.4	A Molscript Diagram of the rEDN Polypeptide Fold	135
Figure 5.5	A Stereo Diagram of Stick Model of the All-Atom Structure of rEDN	137
Figure 5.6	A Stereo Diagram of the Residues of Asp115 to Tyr123 of rEDN and their Packing Interaction with the N-Terminus	140
Figure 5.7	The Interactions Between rEDN and the SO_4(A) and SO_4(B) Anions	142
Figure 5.8	The Intermolecular Packing in the rEDN Crystal	145
Figure 5.9	The Least-Squares Superposition of rEDN, RNase A and Onconase Crystallographic Structures	148
Figure 5.10	The Least-Squares Superposed Active Site Residues and Bound Anions of rEDN and RNase A	151
Figure 5.11	The rEDN-d$(Tp)_3$ Model	152
Figure 5.12	The Nomenclature of the Base, Ribose and Phosphate Moieties of RNA Substrates and Analogs	153

LIST OF ABBREVIATIONS AND SYMBOLS

Å	Angstrom (10^{-10} meters)
B factor	Isotropic thermal motion factor in Å^2 ($B=8\pi\overline{u^2}$)
B_0, B_1, B_2 ...	Base binding subsites
CMP, Cp	Cytidine 3' monophosphate
CpA	Cytidine 3',5' adenylate
CpG	Cytidine 3',5' guanylate
Da	Dalton (1/12 mass of ^{12}C nucleus)
d(CpA)	Deoxycytidine 3',5' deoxyadenylate
d(UpG)	Deoxyuridine 3',5' deoxyguanylate
d(TpTpTpTp)	Deoxythymidylate tetramer
EDN	Eosinophil-derived neurotoxin
F_c	Calculated structure factor amplitude
F_o	Observed structure factor amplitude
I	X-ray scattering intensity
IC_{50}	Concentration required for 50% inhibition
kV	Electrical potential in kilovolts
mA	Electrical current in milliamperes
^{f}Met	Formyl methionine
MIR	Multiple isomorphous replacement
M_r	Molecular weight expressed in Daltons
ONC	Onconase, P-30 Protein
P_0, P_1, P_2 ...	Phosphate binding subsites
P-30 Protein	Onconase
Pyr	Pyroglutamyl, 5-oxo-2-pyrollidine carboxylate
R_0, R_1, R_2 ...	Ribose binding subsites
R factor	Crystallographic residual minimized during refinement ($R=\Sigma\|\|F_o\|-\|F_c\|\|/\Sigma\|F_o\|$)
R_m	Merging R factor ($R_m = \Sigma_{hkl}\Sigma\|I_{ave} - I_{obs}\|/\Sigma\|I_{ave}\|$)
rEDN	Recombinant eosinophil-derived neurotoxin
rms, rmsd	Root mean square, root mean square deviation
RNase	Ribonuclease

rONC	Recombinant Onconase
Up	Uridine 3' monophosphate
V_m	Specific volume in $Å^3$/Dalton
λ	Wavelength of electromagnetic radiation
σ	Standard deviation

CHAPTER 1

INTRODUCTION

Ribonucleases (RNases) can be cytotoxins. In bacteria, fungi, higher plants, amphibians and mammals, there exist ribonucleases that specifically bind to and enter the cytosol of a wide variety of target cells where they degrade RNA species, ultimately killing the target cell. The bacterial cytotoxins, colicin E3 and cloacin DF13, function in inter-strain competition and target ribosomal RNA (rRNA: DeGraaf & Oudega, 1986). The fungal toxin α-sarcin specifically inactivates eucaryotic ribosomes and may be a virulence factor for *Aspergillus* (Chan *et al.*, 1990). The S2 glycoproteins of *Nicotiana alata* function in the prevention of self-pollination and degrades the rRNA of incompatible pollen tubes (Murfett *et al.*, 1994). The cytotoxic RNases of amphibians and mammals are homologous proteins belonging to the pyrimidine-specific RNase superfamily of vertebrates. They include P-30 Protein (Onconase™, ONC) and sialic-acid binding lectins of frog eggs, eosinophil-derived neurotoxin (EDN), eosinophil cationic protein (ECP) and bovine seminal RNase. At present, we do not know how the vertebrate enzymes recognize and enter target cells nor do we know the precise intracellular target of their RNase activity.

The cytotoxic members of the RNase superfamily of vertebrates are of medical interest. The amphibian enzymes have potential therapeutic applications and favorable immunogenic properties (Mikulski et al., unpublished results). Both EDN and ECP have diagnostic applications and play roles in many human pathologies (Youle *et al.*, 1993 and references therein). As a consequence, the ability of these enzymes to specifically bind and enter susceptible cells and their individual ribonucleolytic activities are of interest. In addition, within the RNase superfamily of vertebrates some members are cytotoxic and other are not. Comparison of cytotoxic and non-cytotoxic members of the superfamily offer the possibility of identifying functionally important differences in these enzymes. For

example, which residues contribute to the individual ribonucleolytic specificities of the different enzymes? Which residues or segments of polypeptide chain adopt significantly different conformations in these enzymes? Which residues or segments are responsible for their different biological functions? To address these issues and questions, we determined and analyzed three-dimensional structures of ONC, recombinant ONC (rONC) and recombinant EDN (rEDN). The long term goals of these collaborative projects include the molecular engineering of cytotoxins with enhanced or altered specificity and activity and the development of specific inhibitors of EDN and ECP.

Pyrimidine-specific Ribonuclease Superfamily of Vertebrates

The pyrimidine specific ribonuclease (RNase) superfamily of vertebrates is referred to as the RNase superfamily, pancreatic RNase superfamily and the pyrimidine-specific mammalian alkaline and neutral RNase superfamily in various reviews (Sierakowska & Shugar, 1977; Beintema, 1987; Beintema *et al.*, 1988). The superfamily includes reptilian enzymes (Beintema *et al.*, 1985; Zhao *et al.*, 1994), amphibian enzymes (Ardelt *et al.*, 1991; Titani *et al.*, 1987), avian enzymes (Klenova *et al.*, 1992) and over 40 mammalian members. There are three distinct pyrimidine-specific RNases in the various tissues and body fluids of all tested mammals: the secretory (eg. bovine RNase A) and non-secretory (eg. EDN) RNases defined by Sierakowska & Shugar (1977) and the angiogenins (Strydom *et al.*, 1985). The distinct groups of mammalian enzymes have arisen from gene duplications after the divergence of the mammals and the non-mammals (Beintema *et al.*, 1988) some 150 million years ago and are distant relatives of the reptilian, amphibian and avian enzymes (Beintema *et al.*, 1988).

These enzymes range from 104 amino acids (ONC) to 140 amino acids (bovine brain RNase) in size, have a net positive charge at neutral pH and the mammalian members of the superfamily are glycoproteins. Table 1.1 lists several biochemical properties of representative mammalian members of this superfamily and ONC. Table 1.2 is an amino

Table 1.1

Pyrimidine-specific RNase superfamily of vertebrates

Enzyme†	Amino Acids	Molecular Weight (Da)	Isoelectric Point‡	Function
RNase A	124	13 753	9.2	digestion of RNA in gut
Angiogenin	125	14 615	9.6	angiogenesis
EDN	134	15 463	10.4	host defense
ONC	104	11 845	9.4	unknown (perhaps differentiation)

† RNase A and angiogenin are the bovine enzymes, EDN is isolated from human serum and ONC is from *Rana pipiens* oocytes.

‡ Estimated from the amino acid composition.

acid sequence alignment of representative members of the pyrimidine-specific RNase superfamily of vertebrates. There are 19 residues conserved in the five amino acid sequences of Table 1.2. They include six cysteine residues that form three conserved disulfide bridges, three catalytic residues (RNase A: His12, Lys41, His119) and five residues from two β-strands that form the active site (RNase A: Asn44, Thr45, Phe46, Pro117, Val118), four hydrophobic residues (RNase A: Phe8, Tyr97, Val108), and one residue from a loop (RNase A: Ser77). Table 1.3 lists the percentage amino acid sequence identity between representative mammalian members of the superfamily and ONC. The pair-wise amino acid sequence identity between secretory RNases of different mammals, non-secretory RNases of different mammals and angiogenins of different mammals is typically 70% (Beintema *et al.*, 1988). The relatively low sequence identity and the presence of several large deletions in the Onconase sequence alignment with RNase A originally led us to construct a comparative molecular model of ONC based upon RNase A (Chapter 2). The model addresses the following questions; Can ONC adopt an RNase A-like fold? Is the published amino acid sequence alignment correct? Can Lys9 of ONC replace the conserved Gln (RNase A: Gln14) at the active site of mammalian RNases?

Cytotoxic Ribonucleases of the RNase Superfamily of Vertebrates

The amphibian members of the RNase superfamily include the lectins (*Rana japonica, Rana catesbiana*) and ONC (*Rana pipiens*). These enzymes are present in abundance in oocytes and the early blastomere of frog species and their biological functions are unknown. ONC was initially identified and purified based upon its therapeutic activity in several animal models of cancer (Mikulski *et al.*, 1990). ONC has completed phase I and phase II human clinical trials in the United States as a prospective cancer treatment (Mikulski et al., 1993; unpublished results). It also possesses an anti-HIV activity (Youle *et al.*, 1994) *in vitro* and displays a neurotoxic activity similar to that

Table 1.2

Amino acid sequence alignment of members of the RNase superfamily

	1	*		30	*
Seminal	...kesaaak	FerqHmdsgn	spssssnyCn	lmMccrkmtq	gkCKpvNTFv
RNase A	...ketaaak	FerqHmdsst	saasssnyCn	qmMksrnltk	drCKpvNTFv
EDN	kppqftwaqw	FetqHinmts	q.....qCt	naMqvinnyq	rrCKnqNTFl
Angiogenin	.aqddyryih	FltqHydakp	kgrnd.eyCf	hmMknrrlt.	rpCKpvNTFi
ONC	.....jdwlt	FqkkHitntr	dv.....dCd	niMstnlf..	.hCKdkNTFi
	ααα ααααα		αα ααα		βββββ

	60				90
Seminal	hesladvkav	Csqkkvtckn	gq..tncyqs	kStmritdCr	etgsskyp..
RNase A	hesladvqav	Csqknvackn	gq..tncyqs	ySTmsitdCr	etgsskyp..
EDN	lttfanvvnv	Cgnpnmtcps	nktrknchhs	gSqvplihCn	lttpspqnis
Angiogenin	hgnkndikai	Cedrngqpyr	g....dlris	kSefqitiCk	hkggssrp..
ONC	ysrpepvkai	Ck...qiia.	....sknvlt	tSefylsdCn	vtsrp.....
	β ααααααα α		ββββββ	ββββββββ	

	* 124				
Seminal	nCaYkttqve	khiiVaCggk	.....p	svPVHfdasv	.....
RNase A	nCaYkttqan	khiiVaCegn	.....p	yvPVHfdasv	.....
EDN	nCrYaqtpan	mfyiVaCdnr	dqrrdppqyp	vvPVHldrii	.....
Angiogenin	pCrYgateds	rvivVgCeng	.....	.lPVHfdesf	itprh.....
ONC	.CkYklkkst	nkfcVtCenq	.....	.aPVHfvqvq	sc.....
	βββββββ	βββββββ		βββββ	

A structure based amino acid sequence alignment (one letter code) of representative members of the RNase superfamily . Residues in capitals are conserved among the five sequences and catalytic residues are indicated with an asterix. Residues with α-helical or β -strand conformations in all five structures are indicated by an α (helices) or a β (strands), respectively. The sequence numbers are those of RNase A. The enzymes are bovine seminal RNase, bovine RNase A, human EDN, bovine angiogenin and *Rana pipiens* ONC. The symbol 'j' represents the N-terminal pyroglutamyl of ONC.

Table 1.3

Percent amino acid sequence identity between members of the RNase superfamily

Enzyme†	RNase A	EDN	angiogenin	ONC
RNase A	100	33	32	28
EDN		100	22	27
Angiogenin			100	27
ONC				100

† RNase A and angiogenin are the bovine enzymes, EDN is isolated from human serum and ONC is isolated from *Rana pipiens* oocytes.

of EDN (Newton *et al.*, 1994). The RNases of *Rana japonica* and *Rana catesbiana* were identified as lectins that preferentially agglutinate cancer cells and specifically bind to sialyl-glycoproteins (Titani *et al.*, 1987). The lectins share approximately 53% amino acid sequence identity with ONC and suggested an early model for the mechanism of ONC cytotoxicity. Perhaps, ONC binds to specific cell surface receptors and enters the cell cytosol where it degrades RNA and kills the cell. ONC exhibits saturable binding to the susceptible 9L glioma cells and its affinity is similar to its IC_{50}, 2.5×10^{-7} M. In addition, each of the cytotoxic activities of ONC requires its ribonuclease activity (Wu *et al.*, 1993; Youle *et al.*, 1994; Newton *et al.*, 1994). To further our understanding of the function and enzymatic activity of ONC, its three-dimensional structure has been determined and analyzed (Chapter 3). The structure determination also allows us to critically assess the previously constructed comparative molecular model (Chapter 2). There are several large conformational differences in a comparison of ONC, the comparative molecular model and the non-cytotoxic RNase A structures. Surprisingly, the N-terminal pyroglutamyl residue (Pyr1) folds back against the N-terminal α-helix in the crystal structure and hydrogen bonds to Lys9 at the active site of the enzyme. This suggests Pyr1 has a role in the enzymatic activity of ONC. This is significant as recombinant ONC (rONC) with the N-terminal sequence $^{f}Met_{-1}$, Glu_1, $Asp_2\cdots$ lacks appreciable RNase and cytotoxic activity (Lin et al., 1994). To determine the structural basis of the inactivity of rONC, its structure has been determined and refined at 2.2 Å resolution (Chapter 4).

An additional factor in the cytotoxicity of these enzymes is their sensitivity to the cytosolic ribonuclease inhibitor (RI). The inhibition of pyrimidine specific RNases by the cytosolic RI displays an interesting species specificity. Amphibian RNases are inhibited by the amphibian RI but not by RI of mammalian species and the mammalian RNases are only inhibited by mammalian RIs (Youle *et al.*, 1993). It has been suggested that this cytosolic inhibitor protects cells from RNases of this superfamily (Beintema, 1987).

EDN and ECP share 67% amino acid sequence identity (Gleich et al., 1986) and are

major constituents of the eosinophil granule. Eosinophils release the contents of their granules into the extracellular medium when appropriately stimulated and are a major line of defense against parasites (reviewed in Gleich & Adolphson, 1986). EDN and ECP are clinical indicators of activated eosinophils (Reimert *et al.*, 1993) and are active against parasites in vitro (McLaren *et al.*, 1981). They also play a role in a wide variety of human diseases characterized by excessive eosinophil levels (eosinophilia). These diseases are primarily associated with inflammatory processes and include bronchial asthma (Venge, 1993) and atopic dermatitis (Ott *et al.*, 1994) among many others (Youle et al., 1993 and references therein). EDN and ECP are neurotoxins when injected intracerebrally (Gleich *et al.*, 1986). In animal models the neurotoxicity is expressed as muscle rigidity, incoordination, ataxia and spasmodic paralysis that are consistent with cerebellar deficit. Neuropathological studies show Purkinje cells are selectively sensitive to these RNases (Gordon phenomenon; Durack *et al.*, 1981). As with ONC, the neurotoxicity of EDN and ECP requires its RNase activity (Sorrentino *et al.*, 1992). Precisely how EDN and ECP selectively kill Purkinje cells is still an open question. Unlike the amphibian enzymes, EDN and ECP are strongly inhibited by mammalian RIs (Beintema, 1987).

In parallel with our structural investigations of ONC, the three-dimensional structure of rEDN has been determined and analyzed to examine its function and enzymatic activity (Chapter 5). Of particular interest is a highly polar stretch of residues (Asp115-Tyr123) that are unique to and conserved among the non-secretory members of the superfamily.

Bovine seminal RNase is a unique member of the pyrimidine specific RNase superfamily of vertebrates. It is a covalent homodimer, has significant enzymatic activity towards single and double stranded RNA (ssRNA, dsRNA) and displays anti-spermatogenic and anti-tumor activities that require its RNase activity (for review, D'Alessio, 1993). It shares 81% amino acid sequence identity with the non-cytotoxic, monomeric RNase A. When seminal RNase is prepared as a monomer, it retains its

enzymatic activity towards ssRNA and dsRNA but loses its cytotoxic potential suggesting dimerization is important in its cytotoxic mechanism. Studies showing chemically cross linked dimers of RNase A posses anti-tumor activity (Bartholeyns & Zendergh, 1979) support this interpretation. How dimerization relates to the cytotoxicity of seminal RNase is unclear. Dimerization may facilitate entry into the cell or perhaps its reduced sensitivity to the RNase inhibitor allows it to function as a cytotoxin (Blackburn & Galivanes, 1980). Alternately, the RNase activity of seminal RNase in its dimeric form may lead to its cytotoxicity.

Non-Cytotoxic Members of the RNase Superfamily

RNase A, its major glycosylated form (RNase B) and their mammalian counterparts are secretory RNases that have been intensively studied using biochemical and biophysical techniques (for reviews see Sierakowska & Shugar, 1977; Blackburn & Moore, 1982; Wlodawer, 1985; Beintema, 1987; Beintema *et al.*, 1988). In these reviews and the referenced literature, pancreatic RNase, RNase C, RNase U_I, RNase I, alkaline RNase and neutral RNase are alternate names for the secretory enzymes. The initial translation product of the secretory RNase gene includes a signal peptide and is responsible for the extracellular targeting and the naming of this subfamily of enzymes (Beintema *et al.*, 1988). Secretory RNases of the pancreas are found in considerable quantity in the gut of ruminants and in species with ruminant-like digestion and cecal digestion. In these species the pancreatic RNase content is from 10 and 1000 fold greater than that observed in other herbivores and carnivores. The elevated levels of pancreatic RNase may be the response to the necessity of digesting large amounts of RNA derived from the microflora of the stomach of ruminants (Barnard, 1969). Notably, the extreme differences in the levels of pancreatic RNases among vertebrates are not reflected in the RNase levels of their serum. This implies the existence of an efficient mechanism(s) for regulating serum RNase levels.

Secretory RNases are also expressed at low levels in tissues other than the

pancreas. This would suggest secretory RNases have functions other than digesting RNA. In humans these RNases have a primary structure that is identical to the pancreatic enzyme. In contrast, at least three bovine secretory RNases are known to exist; the pancreatic enzyme, a dimeric form isolated from seminal fluid and the brain RNase that contains 16 additional residues at the C-terminus. These enzymes share 81% and 77% amino acid sequence identity with RNase A (Beintema *et al.*, 1988). At present, only bovine species are known to have these separate secretory RNases.

The non-secretory RNases have been referred to as RNase II, RNase U_S, acid RNase and EDN in the previously mentioned literature. They are somewhat misnamed as non-secretory RNases are found in the urine, serum, cerebrospinal fluid and semen of mammals. EDN is not unique to the eosinophil as its primary sequence is identical to those of non-secretory RNases isolated from human liver, urine, kidney and placenta. EDN and ECP are glycoproteins containing N-linked carbohydrates. Interestingly, the carbohydrate composition of non-secretory RNases differs depending upon the tissue of origin (Yamashita *et al.*, 1986). While the function of the non-secretory enzymes of the various tissues is unknown, EDN retains its neurotoxicity after the enzymatic removal of its N-linked carbohydrate moieties (Gleich *et al.*, 1986). Recently, EDN has been reported to contain a carbohydrate attached to the C2 atom of the Trp7 indole (Hofsteenge *et al.*, 1995). The significance of this novel carbohydrate attachment is not understood.

Angiogenin is not known to be cytotoxic. Instead, it is as an abundantly secreted protein from the human colon adenocarcinoma cell line HT-29 that stimulates the formation of blood vessels (angiogenesis; Strydom *et al.*, 1985). While tumor growth and metastasis may be facilitated by the overexpression and secretion of angiogenin, this is not true in other carcinomas. The mRNA of angiogenin can be detected in a variety of tissues and is predominantly found in liver (Weiner *et al.*, 1987). *In vitro*, angiogenin interacts with an angiogenin-binding protein from endothelial cells (Hu *et al.*, 1991) and elicits a second-messenger response (Heath *et al.*, 1989) that results in angiogenesis. The angiogenic

activity of angiogenin requires its RNase activity (Hallahan *et al.*, 1991) just as ONC and EDN require their RNase activity for their cytotoxicity.

Other than the amphibian RNases, the non-mammalian RNases are not known to be cytotoxins. Instead their relationship to RNase A and the molecular evolution of the RNase superfamily have been examined. The amphibian and avian enzymes form distinct groups of related enzymes that do not conform to mammalian classifications. This is not surprising as the gene duplication events that gave rise to the distinct mammalian enzymes occurred long after the divergence of mammals and non-mammals. Interestingly, the amino acid sequence of turtle ribonuclease indicates it is an angiogenin-like enzyme (Beintema *et al.*, 1988). It shares 43% amino acid sequence identity with angiogenin and less than 30% identity with RNase A and EDN (Table 1.3). This suggests an angiogenin-like gene has been conserved since the divergence of mammals and reptiles and suggests it is the common ancestor of all mammalian pyrimidine-specific RNases.

Ribonucleolytic Mechanism of the Pyrimidine Specific RNases

By definition, these enzymes are those which, at or near neutral pH, cleave RNA endonucleolytically to yield 3'-monoribonucleotides and 3'-oligoribonucleotides ending in Cp or Up via a 2',3'-cyclic phosphate intermediate (Richards & Wyckoff, 1971). The reaction proceeds through a two-step general acid-general base mechanism involving (1) a transesterification in which the 2'-OH nucleophilically attacks the phosphodiester to form the 2',3'-cyclic phosphate intermediate and (2) a hydrolytic step that cleaves the intermediate and yields a product with a terminal 3'-phosphate. The nucleophilic attack by the 2'-OH (step 1) and the solvent molecule (step 2) have both been shown to proceed with an in-line geometry (Usher *et al.*, 1972). Chemical evidence (Crestfield *et al.*, 1963), site-directed mutagenesis (Trautwein *et al.*, 1991), NMR spectroscopy (Rico *et al.*, 1991) and X-ray crystallographic structures of RNase A-inhibitor complexes have agreed that His12, Lys41 and His119 have roles in the catalytic activity of RNase A. Using NMR and X-ray

crystallographic data, a model was proposed (Roberts *et al.*, 1969) in which His12 abstracts the proton of the 2'-oxygen that becomes a nucleophile and attacks the phosphorus, forming a penta-coordinate transition state intermediate. The O(5') leaving group is protonated by His119, the general acid. A salt-bridge between Lys41 and an equatorial oxygen of the phosphate was postulated to stabilize the additional negative charge of the transition state. The refined X-ray crystallographic structures of RNase A in complex with uridine vanadate, a transition-state mimic (Wlodawer *et al.*, 1983) and 3'-cytidine (Zegers *et al.*, 1994) show Lys41 interacts directly with the 2' oxygen as opposed to a phosphate oxygen. This result agrees with charge distribution studies (Davis *et al.*, 1988) and X-ray crystallographic studies on the non-homologous RNase T1 of bacteria that catalyzes an identical reaction. Figure 1.1 is a schematic representation of the reaction catalyzed by members of the pyrimidine specific RNase superfamily of vertebrates.

Recent evidence that imidazole-catalyzed hydrolysis of RNA displays a 'bell-shaped' pH versus rate profile has attracted widespread attention (Anslyn & Breslow, 1989). These results led to a proposed mechanism (triester-like) in which the formation and the breakdown of the pentacovalent phosphorane transition state are catalyzed by different mechanisms. This interpretation differs significantly from the mechanism described above and has been vigorously challenged (Haim, 1992) and defended (Breslow & Xu, 1993) in the literature. Subsequent studies have concluded catalysis by RNase A is not consistent with the triester-like mechanism (Herschlag, 1995) and others have attributed the 'bell-shaped' profile of imidazole-catalyzed hydrolysis of RNA to medium effects (Kirby & Marriott, 1995).

Enzymatic Activity of the Pyrimidine Specific RNases

The enzymatic activity and specificity of the pyrimidine specific RNases have been characterized using diribonucleotides and natural oligoribonucleotide substrates. Unfortunately, the buffers, salts and ionic strengths used to assay the activity of RNase A

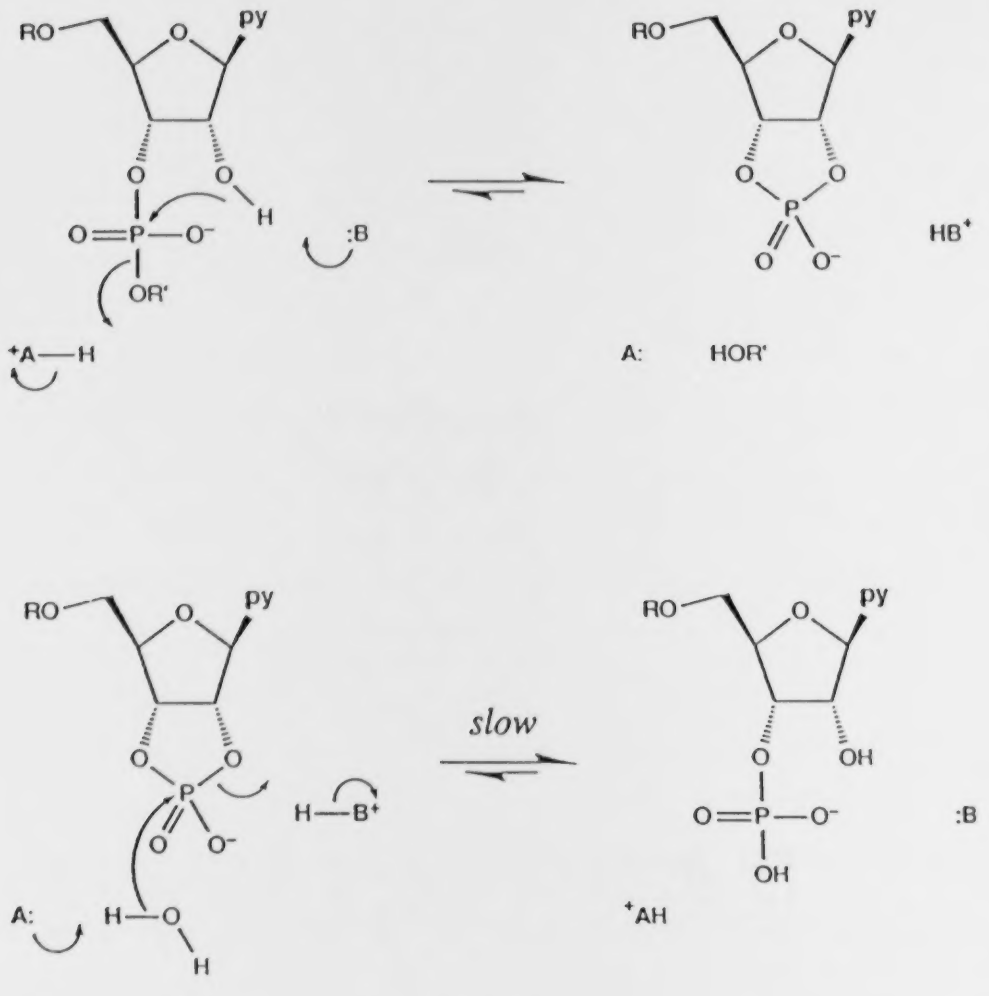


Figure 1.1 The two step ribonucleolytic mechanism shared by members of the pyrimidine (py) superfamily is schematically represented. In the transesterification reaction, the general base abstracts a proton from the O(2') making it nucleophilic and the general acid protonates the leaving group. In the slower hydrolysis step, a solvent molecule is the nucleophile and O(2') is the leaving group. In RNase A, the general acid ($^{+}$A-H) and general base in these reactions are the invariant histidines, H119 and H12, respectively.

affect the experimental results. Despite these difficulties, RNase A and other secretory enzymes are known to be optimally active at approximately pH 8.0 and to generate products ending in Cp or Up. The base, ribose and phosphate moieties of substrates are referred to using the nomenclature of Richards and Wykoff (1971; Figure 1.2). In a recent study of the transesterification reaction, the measured K_m (Michaelis constant) of RNase A for UpX and CpX diribonucleotides (where X is a purine nucleotide) was between 0.3 and 2mM and the k_{cat}/K_m for these substrates was between 0.18×10^5 and $6.0 \times 10^6 M^{-1} sec^{-1}$ (Bond & Vallee, 1990). In UpX and CpX substrates, the U and C are the B_1 nucleoside and X corresponds to the B_2 nucleotide. RNase A shows a slight preference for cytidine in the B_1 position and for adenylate in the B_2 position. Secretory RNases also cleave CpC, UpC, CpU and UpU at much reduced rates and under non-physiological conditions can cleave dsRNA and homopolymers of pA. Table 1.4 list the enzymatic activities of some representative members of the pyrimidine specific RNase superfamily towards various diribonucleotide substrates. The enzymatic activity of EDN is 10 to 100 fold less than of RNase A (Table 1.4) and is optimally active near pH 6.5 (Sierakowska & Shugar, 1977). EDN has a specificity similar to RNase A but with a greater preference for adenylate in the B_2 position. Non-secretory RNases are about 20 times more active against poly(U) substrates than their secretory counterparts, which makes their activity against poly(U) and poly(C) roughly equivalent (Youle *et al.*, 1993). Angiogenin and ONC have pH optima near 6.0. The k_{cat}/K_m of angiogenin and ONC for these substrates is from 10^2 to 10^5 times lower than the comparable values for RNase A and EDN. ONC cleaves UpG and CpG with comparatively high specificity and its specificity is the opposite of the specificity of RNase A. It has a 30 fold preference for uridine in the B_1 position and a 500 fold preference for guanylate in B_2. Angiogenin has a 10 fold preference for cytidine in the B_1 position and has a slight guanylate preference in the B_2 position. Recently, angiogenin and ONC have been shown to preferentially cleave tRNA species in reticulocyte lysates (Saxena *et al.*, 1992, Lin *et al.*, 1994).

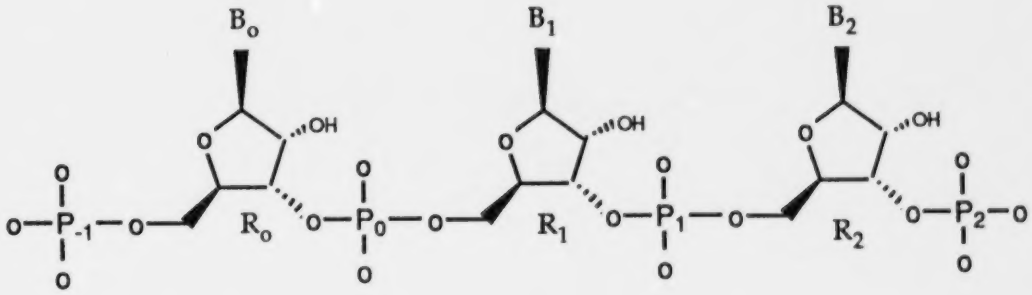


Figure 1.2 A schematic representation of a short RNA substrate. The base, ribose and phosphate moieties are referred to using the nomenclature of Richards and Wykoff, 1971. The scissile phosphodiester bond is P_1-O(5'). Note that the P_1, R_1 and B_1 moieties together are equivalent to a 3'-monophosphate nucleotide. The base, ribose and phosphate numbering increases on the 5' side of the scissile phosphate (P_1) of the RNA substrate and decreases on the 3' side.

Table 1.4

Enzymatic Activity and Specificity of members of the RNase superfamily.

k_{cat}/K_m ($M^{-1}sec^{-1}$)

	UpG	UpA	CpG	CpA	C>p†	U>p†
RNase A[a]	1.8×10^5	4.0×10^6	5.1×10^5	6.0×10^6	7.7×10^3	2.7×10^3
EDN[b]	1.4×10^3	1.2×10^5	1.6×10^3	2.6×10^5	1.1×10^2	ND
Angiogenin[c]	0.4	1.1	4.0	12	0.5	0.02
ONC[d]	4.7×10^3	9.4	1.6×10^2	<2.3	ND	ND

† C>p and U>p are cyclic 2',3' cytidine monophosphate and 2',3' uridine monophosphate, respectively.

[a] Values from Harper & Vallee (1990).

[b] Values from Iwami *et al.* (1981).

[c] Values from Bond & Vallee (1990) and Russo *et al.* (1994).

[d] Values from W. Ardelt and E. Boix, unpublished results.

ND Values have not been determined

Diffraction quality crystals of ONC in complex with a substrate or transition state analogs have not been obtained. To investigate the enzymatic specificity of ONC, a model of ONC has been constructed in complex with UpG, an analog of its preferred substrate. The model allows us to infer the identity of residues that contribute to the B_1 and B_2 specificities of ONC (Chapter 4). In the rEDN structure (Chapter 5), a second sulfate anion binding site has been observed that most likely corresponds to P_{-1} phosphate of extended RNA substrates.

REFERENCES

Anslyn, E. & Breslow, R. (1989) On the mechanism of catalysis by ribonuclease. *J. Am. Chem. Soc.* **111**, 4473-4482.

Ardelt, W.A., Mikulski, S.M. & Shogen, K. (1991) Amino acid sequence of an anti-tumor protein from *Rana pipiens* oocytes and early embryos. *J. Biol. Chem.* **266**, 245-251.

Barnard, E.A. (1969) Biological function of pancreatic ribonuclease. *Nature* **221**, 340-344.

Bartholeyns, J. & Zendergh, A. (1979) *In vitro* and *in vivo* antitumor effect of dimerized ribonuclease A. *Eur. J. Cancer* **15**, 85-90.

Beintema, J.J. (1987) Structure, Properties and Molecular Evolution of Pancreatic-Type Ribonucleases. *Life Chemistry Reports.* **4**, 333-389.

Beintema, J.J., Schuller, C., Irie, M. & Carsana, A. (1988) Molecular Evolution of the Ribonuclease Superfamily. *Prog. Biophys. Molec. Biol.* **51**, 165-192.

Blackburn, P. & Galivanes, J.G. (1980) The role of Lys41 of ribonuclease A in the interaction with RNase inhibitor from placenta. *J. Biol. Chem.* **255**, 10959-10965.

Blackburn, P. & Moore, S. (1982) Pancreatic Ribonuclease. *The Enzymes, 3rd Ed.*, **15**, 317-433.

Bond, M.D & Vallee, B.L. (1990) Replacement of residues 8-22 of angiogenin with 7-21 of RNase A selectively affects protein synthesis inhibition and angiogenesis. *Biochemistry* **29**, 3341-3349.

Breslow, R. & Xu, R. (1993) Quantitative evidence for the mechanism of RNA cleavage by enzyme mimics. *J. Am. Chem. Soc.* **115**, 10705-10713.

Crestfield, A.M., Stein, W.H. & Moore, S. (1963) Alkylation and identification of the

histidine residues at the active site of ribonuclease. *J. Biol. Chem.* **238**, 2413-2420.

Chan, Y., Endo, Y. & Wool, I. (1982) The sequence of nucleotides at the alpha-sarcin cleavage site in rat 28S ribosomal ribonucleic acid. *J. Biol. Chem.* **258**, 12768-70.

D'Alessio, G., DiDonato, A., Parente, A. & Piccoli, R. (1991) Seminal RNase: a unique member of the ribonuclease superfamily. *Trends Biochem. Sci.* **16**, 104-106.

Davis, A., Hall, A. & Williams, A. (1988) Charge description of base catalyzed alcoholysis of aryl phosphodiesters: A ribonuclease model. *J. Am. Chem. Soc.* **110**, 5105-5108.

DeGraaf, F.K. & Oudega, B. (1986) Production and release of cloacin DF13 and related colicins *Current Topics in Microbiology & Immunology* **125**:183-205.

Durack, D.T., Ackerman, S.J., Loegering, D.A., & Gleich, G.J. (1981). Purification of human eosinophil-derived neurotoxin. *Proc. Natl. Acad. Sci. USA* **78**, 5165-5169.

Gleich, G.J. & Adolphson, C.R. (1986) The eosinophilia leukocyte: structure and function. *Adv. Immunol.*, **39**, 177-180.

Gleich, G.J., Loegering, D.A., Bell, M.P., Checkell, J.L., Ackerman, S.J. & McKean, D.J. (1986) Biochemical and functional similarities between human eosinophil-derived neurotoxin and eosinophil cationic protein: Homology with ribonuclease. *Proc. Natl. Acad. Sci.* **83**, 3146-3150.

Haim, A.J. (1992) Imidazole buffer-catalyzed cleavage and isomerization reactions of dinucleotides. *J. Am. Chem. Soc.* **114**, 8384-8388.

Hallahan, T.W., Shapiro, R. & Vallee, B.L. (1991) Dual site model for the organogenic activity of angiogenin. *Proc. Natl. Acad. Sci.* **88**, 2222-2226.

Harper, J., Fox, E., Shapiro, R. & Vallee, B. (1990) Mutagenesis of residues flanking Lys-40 enhances the enzymatic activity and reduces the angiogenic potency of angiogenin. *Biochemistry* **29**, 7297-7302.

Heath, W.F., Moore, F., Bicknell, R. & Vallee, B.L. (1989) Modulation of mitogenic stimuli by angiogenin correlates with *in vitro* phosphatidylinositol bisphosphate synthesis. *Proc. Natl. Acad. Sci. USA* **86**, 2718-2722.

Herschlag, D. (1994) Ribonuclease revisited: Catalysis via the classical general acid-base mechanism or a triester-like mechanism. *J. Am. Chem. Soc.* **116**, 11631-11635.

Hofsteenge, J., Muller, D.R., deBeer, T., Loffler, A., Richter, W.J. & Vliegenthart, J.F.G. (1995) New type of linkage between a carbohydrate and a protein: C-glycosylation of a specific tryptophan residue in human RNase U_S *Biochemistry* **33**, 13524-13530.

Hu, G., Chang, S., Riordan, J.F. & Vallee, B.L. (1991) An angiogenin-binding protein from endothelial cells. *Proc. Natl. Acad. Sci. USA* **88**, 2227-2231.

Iwami, M., Kunihiro, M., Ohgi, K. & Irie, M. (1981) Purification and properties of human urine ribonuclease. *J. Biochem.* **89**, 1005-1016.

Kartha, G., Bello, J. & Harker, D. (1967) The tertiary structure of ribonuclease. *Nature* **213**. 862-865.

Kirby, A.J. & Marriott, R.E. (1995) Mechanism of RNA cleavage by imidazole. *J. Am. Chem. Soc.* **117**, 833-834.

Klenova, E.M., Botezato, I., Laudet, V., Goodwin, G.H., Wallace, J.C. & Lobanenkov, V.V. (1992) Isolation of a cDNA encoding an RNase related gene highly expressed in chicken bone marrow cells. *Bioch. Biophys. Res. Commun.* **185**, 231-239.

Lin, J.J., Newton, D.L., Mikulski, S.M., Kung, H.F., Youle, R.J. & Ryback, S.M. (1994) Characterization of the mechanism of cellular and cell free protein synthesis inhibition by an anti-tumor ribonuclease. *Bioch. Biophys. Res. Com.*. **204**, 156-162.

McLaren, D.J., McKean, J.R., Olsson, I., Venge, P. & Kay, A. (1981) Morphological studies on the killing of schistosomula of *Schistosomula mansoni* by human

eosinophil and neutrophil cationic protein *in vitro*. *Parasitol. Immunol.*, **3**, 359-365

Mikulski, S.M., Ardelt, W., Shogen, K., Bernstein, E.H., and Menduke, H. (1990). Striking increase of survival of mice bearing M109 Madison carcinoma treated with a novel protein from amphibian embryos. *J. Natl. Cancer Inst.* **182**, 151-153

Murfett, J., Atherton, T.L., Mou, B., Gasser, C.S. & McClure, B.A. (1994) S-RNase expressed in transgenic Nicotiana causes S-allele-specific pollen rejection. *Nature.* **367**:563-566.

Newton, D.L., Walbridge, S., Mikulski, S.M., Ardelt, W., Shogen, K., Ackerman, S.J., Ryback, S.M. & Youle, R.J. (1994) Toxicity of an antitumor ribonuclease to Purkinje Neurons. *J. Neuroscience* **14**, 538-544.

Ott, N.L., Gleich, G.J., Peterson, E.A., Fujisawa, T., Sur, S. & Leiferman, K.M. (1994) Assessment of eosinophil and neutrophil participation in atopic dermatitis: comparison with the IgE-mediated late-phase reaction *J. Allergy Clinical Immunol.* **94**, 120-128.

Reimert, C.M., Poulsen, L.K., Bindslev-Jensen, C., Kharazmi, A. & Bendtzen, K. (1993) Measurement of eosinophil cationic protein and eosinophil protein X/eosinophil-derived neurotoxin. *J. Immunol. Methods*, **166**, 183-190.

Richards, F.M. & Wykoff, H.W. (1973). Atlas of molecular structures in biology (Phillips, D.C. & Richards, F.M., eds.). vol. 1, RNase A, Clarendon Press, Oxford.

Rico, M., Santoro, J. Gonzalez, C., Bruix, M., Neira, J., Nieto, J. & Herranz, J. (1991) 3D structure of bovine pancreatic ribonuclease A in aqueous solution. *J. Biomol. NMR* **1**, 283-298.

Roberts, G.C, Dennis, E.A., Meadows, D.H., Cohen, J.S. & Jardetzky, O. (1969) The mechanism of action of ribonuclease. *Proc. Natl. Acad. Sci. USA* **62**, 1151-1158.

Russo, N, Shapiro, R., Acharya, K.R., Riordan, J., & Vallee, B.L. (1994) Role of glutamine-117 in the ribonucleolytic activity of human angiogenin. *Proc. Natl. Acad.*

Sci. USA **91**, 2920-2924.

Saxena, S.K., Ryback, S.M., Davey, R.T., Youle, R.J. & Ackerman, E.J. (1992) A cytotoxic tRNA specific ribonuclease of the RNase A superfamily. *J. Biol. Chem.* **267**, 21982-21986.

Sierakowska, H. & Shugar, D. (1977) Mammalian nucleolytic enzymes. *Prog. Nucleic Acid Res. molec. Biol.* 20,59-130.

Strydom, D.J., Fett, J.W., Lobb, R.R., Alderman, E.M., Bethune, J.L., Riordan, J.F. & Vallee, B.L. (1985) The amino acid sequence of human tumor derived angiogenin. *Biochemistry* **24**, 5486-5494.

Sorrentino, S., Glitz, D.G., Hamann, K.J., Loegering, D.A., Checkel, K.L. & Gleich, G.J. (1992) Eosinophil-derived neurotoxin and human liver ribonuclease. *J. Biol. Chem.*, **267**, 14859-14865.

Strydom, D.J., Fett, J.W., Lobb, R.R., Alderman, E.M., Bethune, J.L., Riordan, J.F. & Vallee, B.L. (1985) Amino acid sequence of human tumor derived angiogenin. *Biochemistry* **24**, 5486-5492.

Titani, K., Takio, K., Kuwada, M., Nitta, K., Sakakibara, F., Kawauchi, H., Takayanagi, G. & Hakomori, S. (1987) Amino acid sequence of sialic acid binding lectin from frog (*Rana catesbiana)* eggs. *Biochemistry* **26**,2189-2194.

Trautwein, K., Holliger, P., Stackhouse, J. & Benner, S.A. (1991) Site-directed mutagenesis of bovine pancreatic ribonuclease: Lysine-41 and Asp-121. *FEBS Letters*, **281**, 275-277.

Usher, D.A., Erenrich, E.S., and Ekstein, F. (1972). Geometry of the first step in the action of ribonuclease-A. *Proc. Natl. Acad. Sci. (USA)* **69**, 115-118

Venge, P. (1993) Eosinophil activity in bronchial asthma. *Allergy Proc.*, **15**,139-141.

Weiner, H.L., Weiner, L.H. & Swain, J.L. (1987) Tissue distribution and developmental expression of messenger RNA encoding angiogenin. *Science* **237**, 280-282.

Wlodawer A, Miller M, Sjolin L. 1983. Active site of RNase: neutron diffraction study of a complex with uridine vanadate, a transition state analog. *Proc. Natl. Acad. Sci. USA 80:*3628-3631.

Wlodawer, A. (1985) Nucleic Acids and Interactive Proteins. (Jurnak & McPherson, Eds.). *Biological Macromolecules and Assemblies* . **Vol. 2**, 395-439.

Wu, Y., Mikulski, S.M., Ardelt, W., Ryback, S.M. & Youle, R.J. (1993). A cytotoxic ribonucleases: Study of the mechanism of Onconase cytotoxicity. *J. Biol. Chem.* **268**, 10686-10693.

Yamashita, K., Hitoi, A., Irie, M. & Kobuta, A. (1986) Fractionation by lectin affinity chromatography indicates that the glycosylation of most ribonucleases in human viscera and body fluids is organ specific. *Arch. Biochem. Biophys.* **250**, 263

Youle, R.J., Newton, D., Wu, Y., Gadino, M. & Ryback, S.M. (1993) Cytotoxic ribonucleases and chimeras in cancer therapy. *Critical Rev. Therapeutic Drug Carrier Systems* **10**, 1-28.

Youle, R.J., Wu, Y., Mikulski, S.M., Shogen, K., Hamilton, R., Newton, D., D'Alessio, G. & Gravell, M. (1994) RNase inhibition of HIV infection of H9 cells. *Proc. Natl. Acad. Sci. USA* **91**, 6012-6016.

Zegers, I., Maes, D., Thi, M.H.D., Wyns, L., Poortmans, F., Palmer, R. (1994) The structure of RNase A complexed with 3'-CMP and d(CpA): Active site conformation and conserved water molecules. Protein Science **3**:2322-2339.

Zhao, W., Beintema, J.J. & Hofsteenge, J. (1994) The amino acid sequence of iguana pancreatic ribonuclease. *Eur. J. Biochem.* **219**, 641-646.

CHAPTER 2

COMPARATIVE MOLECULAR MODELING AND CRYSTALLIZATION OF P-30 PROTEIN: A NOVEL ANTI-TUMOR PROTEIN OF *RANA PIPIENS* OOCYTES AND EARLY EMBRYOS†

INTRODUCTION

P-30 Protein (Onconase)‡ is an abundant, basic protein that is isolated from the oocytes or early embryos of *Rana pipiens* (1). It exhibits a specific anti-tumor activity against several human tumor cell lines *in vitro* including notoriously resistant solid tumors (1,2). The P-30 Protein (hereafter referred to as P-30) is also effective against the M109 Madison carcinoma in mice (3). Treatment of susceptible tissue cultures with P-30 results in the accumulation of cells arrested in G_1 phase of the cell cycle and cytotoxicity. Subsequent examination of those cells arrested in G_1 phase reveals heterogeneous total RNA contents, with many cells exhibiting very low levels of RNA (1). This suggests P-30 is able to preferentially enter susceptible cells by an undetermined mechanism. P-30 is currently undergoing phase II human clinical trials in the United States as a prospective solid tumor treatment.

The complete amino acid sequence of P-30 has recently been published (4). The protein consists of 104 amino acids including four disulfide bridges and a blocked N-terminus (pyroglutamyl). The calculated molecular weight is 11,837 Da when basic amino acid side chains are assumed to be fully protonated. More interestingly, it has been shown that P-30 is a member of the pancreatic RNase superfamily of enzymes (4). Currently, P-30's ribonucleolytic activity has been confirmed, however, its catalytic activity is low in

---

† A version of this chapter has been published. Mosimann, S.C., Johns, K.L., Ardelt, W., Mikulski, S.M., Shogen, K. & James, M.N.G. (1992). Comparative molecular modeling and crystallization of P-30 protein: a novel anti-tumor protein of *Rana pipiens* oocytes and early embryos. *Prot. Structure Function and Genetics* **14**, 392-400.

‡ Onconase is a trademark of the Alfacell Corporation.

comparison with other members of this family. It is intriguing that P-30's anti-tumor activity is completely inhibited by chemical modification of suspected catalytic residues (4). This suggests that P-30's ribonucleolytic activity is somehow required for its anti-tumor effect.

In light of P-30's exhibited anti-tumor activity, and in the absence of three dimensional structural information, a comparative molecular model (5-7) has been constructed with a view towards understanding the specificity of P-30. The pancreatic RNases have been extensively studied by physiochemical methods (8) and we are fortunate to have several refined, high resolution ($\leq$ 2.0 Å) structures of bovine pancreatic RNase A (9-11). In addition, molecular structures of RNase S (an active form of RNase A resulting from subtilisin cleavage between Ala20 and Ser21) are available (12-13). In this work, RNase A's crystallographic coordinates serve as a template for the comparative molecular modeling of P-30.

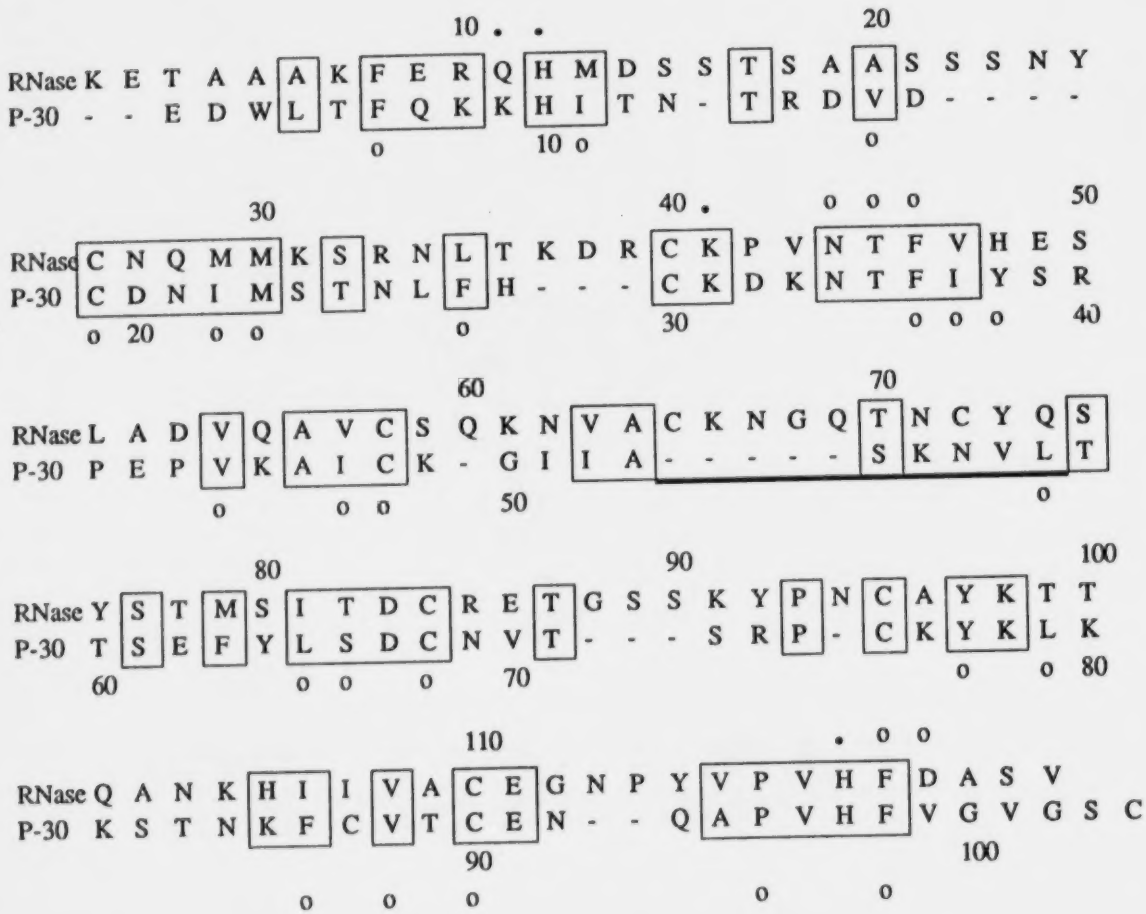
RESULTS

Alignment of P-30 and RNase A sequence — One of the most critical steps in the construction of a comparative molecular model is the alignment of the amino acid sequences, including the placement of both insertions and deletions. The current axiom for modeling homologous structures is attributed to J. Greer; if the alignment is wrong, the structure is guaranteed to be wrong (14,15).

In the present case, P-30 is considerably smaller (104 residues) than bovine pancreatic RNase A (124 residues). The sequence alignment we have deduced is shown in Table 2.1. It results from a two step process. First, the use of automated alignment programs (16-18), and subsequently, a revision of the alignment based on inspection of the P-30 sequence fitted into the RNase A structure. The present alignment makes use of three conserved disulfide bridges, the three catalytic residues and their flanking residues to align the bulk of the sequence. It should be pointed out that one of the disulfides of RNase A

Table 2.1

Complete amino acid sequence alignment of pancreatic RNase A and P-30 Protein


RNase K E T A A A K F E R Q H M D S S T S A A S S S N Y
P-30 - - E D W L T F Q K K H I T N - T R D V D - - - -
RNase C N Q M M K S R N L T K D R C K P V N T F V H E S
P-30 C D N I M S T N L F H - - - C K D K N T F I Y S R
RNase L A D V Q A V C S Q K N V A C K N G Q T N C Y Q S
P-30 P E P V K A I C K - G I I A - - - - - S K N V L T
RNase Y S T M S I T D C R E T G S S K Y P N C A Y K T T
P-30 T S E F Y L S D C N V T - - - S R P - C K Y K L K
RNase Q A N K H I I V A C E G N P Y V P V H F D A S V
P-30 K S T N K F C V T C E N - - Q A P V H F V G V G S C

Amino acid homology between RNase A and P-30 Protein are indicated as boxed residues*. The underscored region represents a revision of the original sequence alignment based on the inspection of the RNase A structure. Open circles above residues indicate constituents of the active site while closed circles above the sequence represent catalytic residues. Open circles below the sequence alignment indicate residues contributing to the hydrophobic core.

*28% identity and 48% homology.

(Cys65 to Cys72) has no counterpart in P-30, whereas the fourth disulfide bridge of P-30 (Cys87 to Cys104) involves an inserted residue, the C terminal cysteine. An example of sequence alignment revision based on inspection of the RNase A structure is underlined in Table 2.1. In this region P-30 contains a five residue deletion relative to the RNase A sequence Cys65 to Gln74. Various automated alignment programs (16-18) position the deletion at RNase A residues Thr70 to Gln74, as a result of the strong homology at Cys65 to Asn67 (RNase A: Cys Lys Asn, P-30: Ser Lys Asn). However, in light of the structure of RNase A, it is clear that residues Cys65 to Gln69 form part of an exposed loop (bracketed by the Cys65-Cys72 disulfide bridge), whereas residues Thr70 to Gln74 contribute to one of the central strands of a four stranded anti-parallel β-sheet that constitutes part of RNase A's core. Figures 2.1a and 2.1b illustrate the consequences of positioning the deletion at Cys65 to Gln69 and Thr70 to Gln74 respectively. Since deletion of the Thr70 to Gln74 β-strand is likely to change RNase A's overall fold, we have chosen to position this deletion at residues Cys65 to Gln69.

The resulting alignment of P-30 with RNase A has 28% (29/104) sequence identity. When conservative amino acid replacements are included, the sequence alignment has 48% (50/104) similarity (Ala = Val = Leu = Ile = Met = Phe = Tyr, Glu = Gln = Asp = Asn, His = Arg = Lys, Ser = Thr). A complete accounting of the sequences deleted in the alignment of P-30 with RNase A is shown in Figure 2.2, where it can be seen that deleted sequences map mainly to exposed loops on the surface of RNase A. As a result, the alignment is expected to conserve the secondary structural elements of RNase A and many features of its hydrophobic core, in addition to its active site.

Model building and energy minimization — The program MUTATE (Randy Read, unpublished) was used to provide atomic coordinates for the P-30 sequence. In this process the amino-acid side chains of RNase A, were replaced by the side chains of P-30 according to the sequence alignment given in Table 2.1. In cases where small residues in RNase A were replaced by larger side chains in P-30, the extra atoms of the side chain were

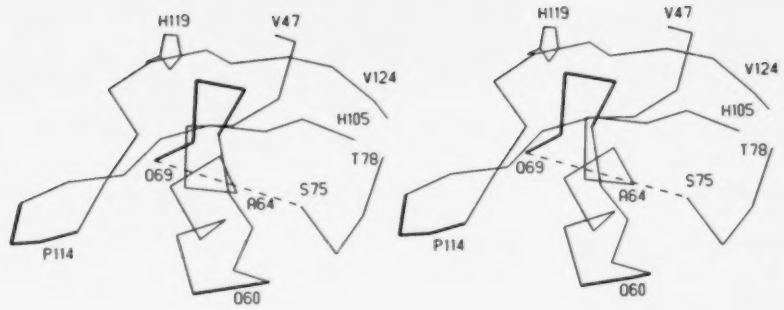


Figure 2.1a An example of sequence alignment revision based upon inspection of RNase A is shown. The dashed lines indicate regions that are deleted and the thick lines represents regions that are retained in two potential sequence alignments. Automated alignment programs (16-18) predict that RNase A residues Thr70-Gln74 are deleted relative to P-30 Protein's sequence.

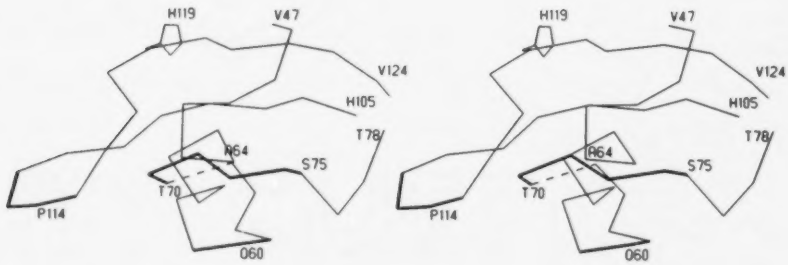


Figure 2.1b: Again, the dashed lines indicate regions that are deleted and the thick lines represents regions that are retained in two potential sequence alignments. In this alignment, Cys65-Gln69 are deleted. This has the advantage of preserving the antiparallel β-sheet (which forms part of RNase A's core) and leaving the ends of the deletion near to one another. The model of P-30 Protein has made use of the alignment depicted in b.

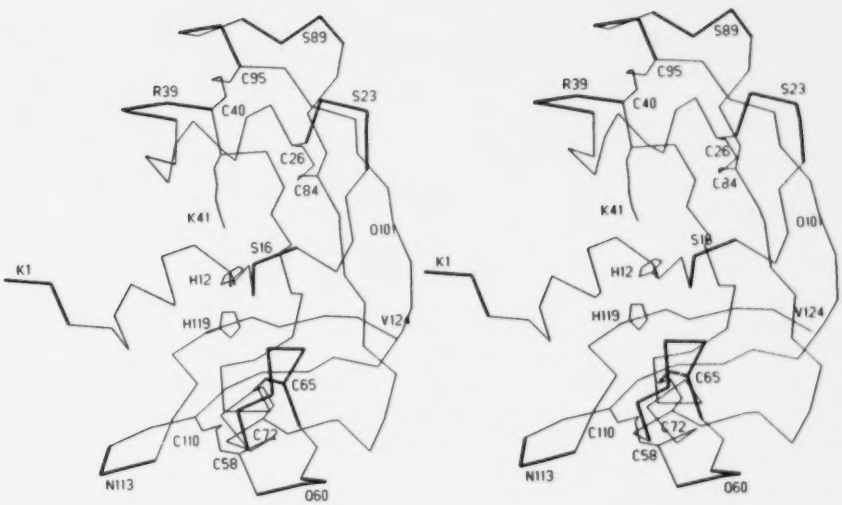


Figure 2.2: A C^{α} representation of RNase A is presented with deletions (relative to P-30 Protein alignment in Table 2.1) indicated in thick lines and disulfide bridges shown as dashed lines. Note that the deletions occur primarily in exposed loops of the RNase A structure.

positioned according to expected stereochemistry (19,20). The inevitable too close van der Waal's contacts that arise from such amino acid replacements were relieved using the interactive graphics program TOM (21,22). As an example, the too close contacts that result from the replacement of Ala5 (RNase A) by Trp3 (P-30) are shown in Figure 2.3. Figure 2.3 also shows the identical region after graphical intervention. In this instance, rotations about the torsion angles χ_1 and χ_2 of Trp3, facilitate the location of a low energy conformation that is free of steric clash. A noteworthy consequence of this reorientation, is the 3.10 Å separation between $N^{\varepsilon 1}$ of Trp3 and the carbonyl oxygen of Glu91 suggesting a possible hydrogen bond interaction.

Once all non-hydrogen contacts less than an arbitrarily chosen distance of 2.6 Å were relieved in this manner, the residues at each end of a deletion and the residues adjacent to them in the primary sequence were subjected to 200 cycles of stereochemical regularization using J. Hermann's routine within TOM (21,22). At this point, the additional C terminal Ser103 and Cys104 residues were added to the structure, and the disulfide bridge involving the C-terminal cysteine (Cys87 to Cys104) was constructed, again using J. Hermann's regularization routine within TOM. The P-30 model was then subjected to energy minimization using the CHARMm program within QUANTA (23). The relatively robust steepest descents minimization (25 cycles) procedure was used to tidy up unacceptably close contacts and poor stereochemistry in the model. This was followed by conjugate gradient minimization until convergence, at which point, the N-terminal glutamate was cyclized and manipulation of the model was complete.

As all energy minimization was done in the absence of solvent and ions, some accounting of potential artifacts is required. To this end, the crystallographic coordinates of the RNase A structure were also subjected to energy minimization, under conditions identical to those used on the P-30 model coordinates. In each case, energy minimization was performed with model polar hydrogens, and parameters were invoked which assign electrostatic neutrality to potentially charged residues in an effort to account for the

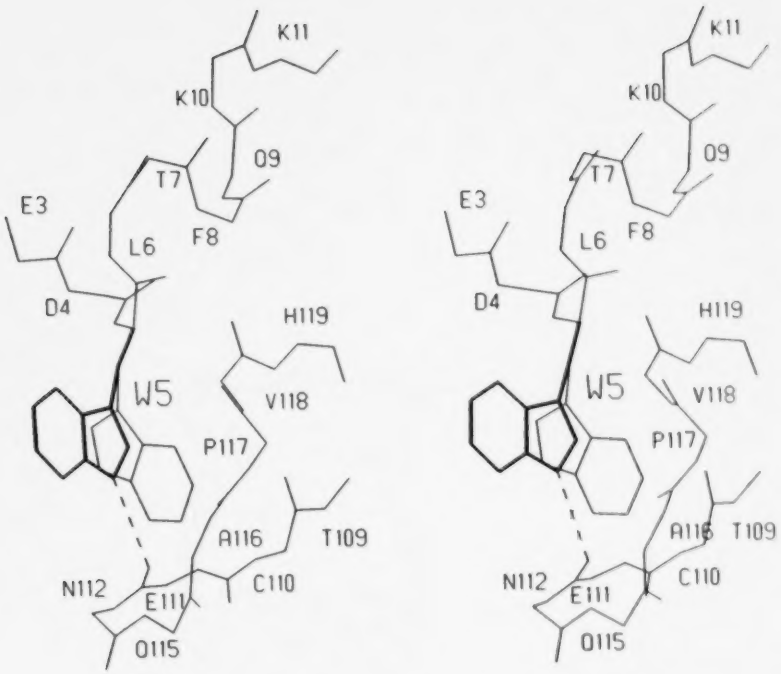


Figure 2.3: An extreme example of unacceptable close contacts introduced by the replacement of Ala5 (RNase A) with Trp (P-30 Protein) and its relief through graphical intervention. The RNase A numbering scheme is used in the figure and the thick lines correspond to the reoriented Trp. In this instance, the rotation of χ_1 and χ_2 torsion angles of Trp has relieved close contacts and indicates the potential formation of a H-hydrogen bond (shown as a dashed line) between Trp $N^{\varepsilon 2}$ and a mainchain carbonyl oxygen.

lack of solvent. Finally, it should be stated that the model can only be considered tentative.

RNase A and energy minimized RNase A — The root mean square (rms) deviation in atomic coordinates for the minimized structure of RNase A and the model of P-30 from the crystallographic structure of RNase A are listed in Table 2.2. The rms deviation of 0.59 Å between all atoms of the crystallographic RNase A structure and the energy minimized RNase A structure, indicates that gross structural perturbations have not been introduced by the energy minimization procedures adopted. As an indicator, each of the secondary structural elements of RNase A remain intact throughout energy minimization procedure. This is confirmed visually in Figure 2.4, where the RNase A structures (C^{α} and selected residues), before and after the minimization process have been superposed. It is noteworthy, that in Table 2.2, the all-atom, active site only superposition of RNase A before and after energy minimization actually has an rms deviation greater than that of the remainder of the structure. In fact, the side chains of the catalytic residues His119 and Lys41 (shown in Figure 2.4) undergo some of the largest movements between structures. Under normal circumstances these results would be distressing. However, the crystallographic structure of RNase A contains a bound phosphate anion that interacts non-covalently with each of the three catalytic residues; His12, Lys41 and His119. Thus, it is presumed that the absence of the phosphate anion during energy minimization has led to the significant reorientation of His119 and Lys41 in the minimized RNase A structure. Overall, the series of small differences between the two structures is expected to reflect errors in both the crystallographic coordinates and probably to a greater extent, assumptions inherent in current energy minimization strategies.

When constructing a comparative molecular model, energy minimization procedures serve to relieve too close van der Waal's contacts and to ensure chemically reasonable structures. Given the agreement between the crystallographic RNase A structure and the energy minimized RNase A structure, it is expected that this energy minimization strategy will yield meaningful results when applied to the P-30 model coordinates.

Table 2.2

*Rms deviation in atomic coordinates for minimized RNase A and the P-30 Protein model**

	all atoms (Å)	mainchain (Å)$^{\Delta}$	all active site (Å)°
RNase A (crystal) vs. RNase A (minimized)	0.59	0.42	0.63
P-30 Protein model vs. RNase A (minimized)	1.38	1.29	0.52

* Comparisons between atoms in common in accordance with alignment in Table 2.1.

$^{\Delta}$ Mainchain atoms include N, C^{α}, C and O.

° The active site includes all atoms of the catalytic triad (His12, Lys41, His119) and 6 other residues that contribute to the substrate binding cleft (Gln11, Asn44, Thr45, Phe46, Phe120 and Asp121).

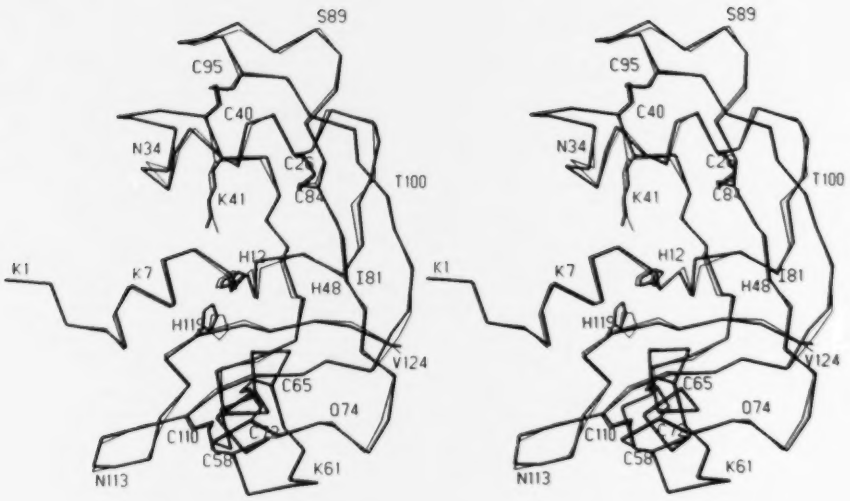


Figure 2.4: Superposition of RNase A crystallographic structure before and after energy minimization. Only C^{α} atoms, catalytic residues and disulfide bridges are shown. The minimized RNase A structure is shown in heavy lines and disulfide bridges are shown as dashed lines.

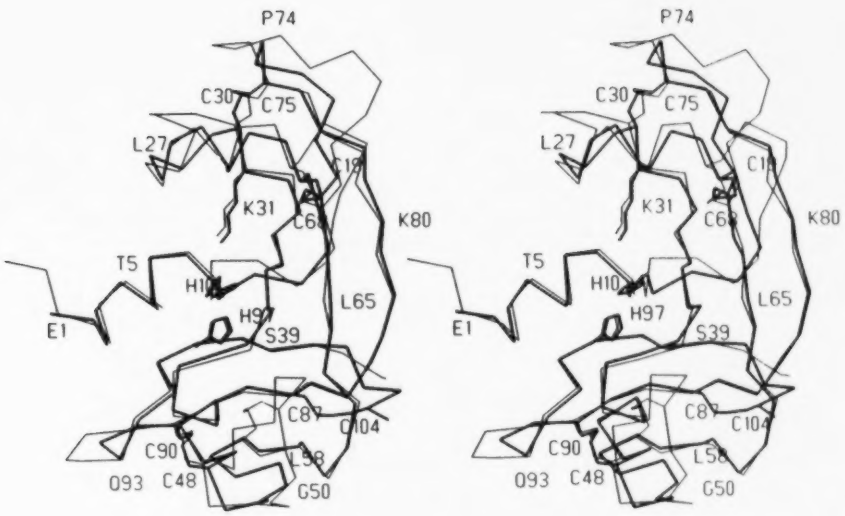


Figure 2.5: Superposition of minimized RNase A and the P-30 Protein model. Again, C^{α} atoms, catalytic residues and disulfide bridges are presented. The P-30 Protein model is represented by the heavy lines in this figure and disulfide bridges are shown as dashed lines.

P-30 comparative molecular model — The rms deviations for the several superpositions of P-30 and minimized RNase A are presented in Table 2.2. Figure 2.5 shows a superposition of the catalytic residues, disulfide bridges and C^{α} atoms of P-30 and the minimized RNase A.

Architecturally, RNase A consists of three regions; an N terminal helix (~ 15 residues) and two larger β sheet containing elements (~ 45 residues) that are separated by a hinge of 3 peptide strands. The two β sheet containing elements are related by approximate 2 fold symmetry and each of these regions contributes a catalytic residue to the active site. In more detail, the N-terminal helix of P-30 (S peptide of RNase S) provides Lys9 and His10 (Gln11 and His12 in RNase A) to the active site and makes contacts with both of the remaining regions. The active site Lys31 and Lys33-Thr35 of P-30 and the corresponding residues of RNase A (Lys41 and Val43-Lys45) are supplied by the large region that is composed of a relatively long three-stranded antiparallel β-sheet and a short helix that packs against the β sheet on the far side of the active site. The remaining element contributes His97-Val99 (His119-Asp121 in RNase A) and is formed by a shorter four-stranded antiparallel β-sheet. This domain also contains a short helix that packs against the β sheet on the side furthest from the active site.

As previously indicated, the sequences in RNase A that are deleted in the P-30 model, primarily involve exposed surface loops. Excluding the two residues deleted from the N terminus of P-30, there are five deletions involving more than one residue (see Table 2.1). In each of these cases, the secondary structural elements that bracket the deleted sequence are retained in the comparative molecular model of P-30. This conservation of secondary structural elements is related to the preservation of both identical and equivalent tertiary interactions that form the hydrophobic core of RNase A. In Table 2.1 it can be seen that the alignment of hydrophobic core residues is 50% (12/24) identical and 79% (19/24) homologous.

Of the five deletions referred to, three occur at a distance from the active site

(Ser22-Tyr25, Gly88-Ser90, Asn113-Pro114) of RNase A. The two remaining deletions (Lys37-Arg39, Cys65-Gln69) map to the periphery of the RNase A active site. Of these the Lys37-Arg39 deletion is only one residue removed from the catalytic Lys41. Despite this, the Lys31 coordinates (P-30) can remain essentially unchanged when compared to Lys41 of the minimized RNase A structures. It is notable that the intervening residue (Cys40 of RNase A; Cys30 in P-30) is involved in a conserved disulfide bridge that presumably contributes to the correct positioning of the catalytic Lys41 in RNase A (Lys31 in P-30). With reference to the deduced mechanism for pancreatic RNases, the catalytic Lys interacts with the pentacovalent transition state. The other deletion, Cys65-Gln69, has been previously discussed and the chosen alignment is somewhat justified by the involvement of Ser54-Leu58 in the four-stranded β-sheet of the P-30 model (Thr70-Gln74 of RNase A).

The sole insertion in the alignment of P-30 and RNase A sequences involves the two C-terminal residues, Ser103 and Cys104. The C-terminal Cys104 participates in a disulfide bridge with Cys87, that is unique to P-30. These additional residues are peripheral to the active site and form a turn on the surface of the P-30 model, immediately following the active site residues His97-Val99.

With reference to the active site residues mentioned previously and defined again in Table 2.2, P-30 only differs from RNase A in three of nine positions. The increased conservation of active site residues is reflected in the all atom, active site rms deviation, which is only 0.52 Å. When comparing the active site differences between P-30 and RNase A, the respective replacements are Lys9, Lys33 and Val99 (P-30) for Gln11, Val43 and Asp121 (RNase A). In an effort to evaluate these amino acid replacements in terms of the specificity of P-30, various crystallographic structures of RNase S- (8,24-26) and RNase A-inhibitor complexes (27,28) have been examined. In general, the results of the inhibitor structures are internally consistent and supportive of earlier chemical studies (29-30). Unfortunately, these studies have been hampered by the lack of available

coordinates for RNase-inhibitor complexes.

The specificity of RNase A is such that the phosphate at the active site has a 3' linkage to the pyrimidine ribonucleotide and a scissile 5' linkage to the remaining ribonucleotide. In native and inhibitor structures of RNase A, $N^{\varepsilon 2}$ of Gln11 donates a hydrogen bond to the active site phosphate. In the P-30 model, Lys9 has the potential to form a similar hydrogen bond. Despite the increased size (1 C-C bond) of Lys9, there are few steric constraints imposed by neighboring residues, and thus it is possible that this contact is retained in both P-30 and RNase A.

The contribution of Val43 to the RNase A active site is primarily through its mainchain atoms. As such, the non-conservative Val43 to Lys33 replacement present in P-30 probably will not alter substrate binding significantly. Alternatively, the additional positive charge introduced by this substitution could affect the substrate specificity of P-30.

The greatest active site differences between RNase A and P-30 involve the Asp121 to Val99 change in conjunction with the deletion of the Cys65-Gln69 loop. In the several RNase structures, Asp121 makes side chain contacts with the Cys65-Gln69 containing loop, and the active site His97 side chain. These contacts remain unchanged in structures of RNase A-inhibitor complexes. However, one or both of Gln69 and Asn71 contribute side chain hydrogen bonds to the 3' base (purine or pyrimidine) in diribonucleotide inhibitor-RNase A complexes (24-26). In the P-30 model, the Cys65-Gln69 region of RNase has been deleted and thus P-30 most likely will not form the analogous interactions. It is noteworthy, that this active site difference involves the 3' ribonucleotide binding site, since RNase A does not show a rigid specificity at this site. To expand on this point, it has been shown biochemically, that RNase A activity towards diribonucleotides, varies over two orders of magnitude in response to various 3' ribonucleotides (31). This tolerance of RNase A for 3' ribonucleotides would suggest that other interactions are of greater importance in substrate binding and catalysis.

In summary the comparative molecular model of P-30 has been constructed with the

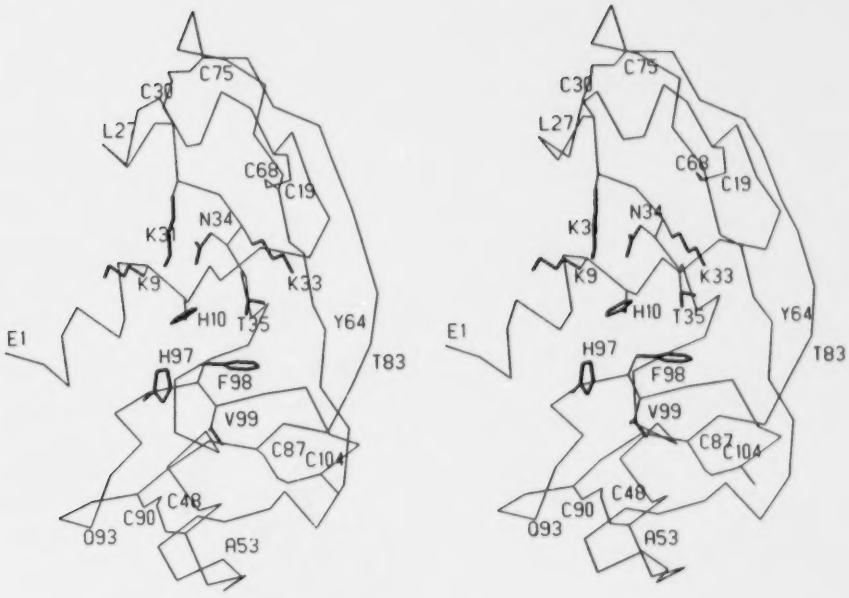


Figure 2.6: The tentative P-30 Protein model. Active site residues are included as well as catalytic residues and disulfide bonds. The active site and catalytic residues are shown in thick lines and the disulfide bridges are indicated by dashed lines.

aim of understanding some of the underlying features of its specificity. In comparison with RNase A, the P-30 model has three significant active site differences which are shown in Figure 2.6. First, Lys9 of P-30 replaces Gln11 of RNase A which donates an important hydrogen bond to the active site phosphate. Second, there are topological changes at the periphery of the active site that result from an insertion and deletions. Since these regions of RNase enzymes may function in the binding of oligoribonucleotides, they may contribute to the specificity of P-30. Finally, the absence of the Cys65-Gln69 (RNase A) loop in P-30, indicates that P-30 probably cannot bind the 3' ribonucleotide in a fashion similar to RNase A. As P-30's specificity with respect to various substrates is currently unknown, it will be interesting to note the similarities and differences between the specificities of P-30 and RNase A as more data become available.

Crystallization — Recently, our lab has grown two separate forms of P-30 crystals, orthorhombic and hexagonal. Of these, the orthorhombic form has been characterized by X-ray diffraction techniques.

Crystals of P-30 were grown by vapor diffusion using a hanging drop. Typically, 2 μL of 15 mg/mL P-30 (in H_2O) are mixed with 2 μL of reservoir solution (50% saturated $(NH_4)_2SO_4$), 10 mM acetate pH 4.5) and allowed to equilibrate with 1 mL of reservoir solution. Crystals grow as rosettes in 3-7 days and achieve maximum size after approximately 1 month. Single crystals can be broken off the rosettes for subsequent diffraction experiments. The crystals have unit cell dimensions of $a = 40.76$, $b = 69.77$, and $c = 32.54$ Å and belong to the space group $P2_12_12_1$. We are currently optimizing the crystallization conditions in order to obtain large single crystals suitable for high resolution diffraction experiments. It is hoped that this comparative molecular model will contribute to the solution of the three dimensional structure of P-30, when satisfactory crystals are obtained.

REFERENCES

1. Darzynkiewicz, Z., Carter, S.P., Mikulski, S.M., Ardelt, W.J., and Shogen, K. (1988). Cytostatic and cytotoxic effects of Pannon (P-30 Protein), a novel anticancer agent. *Cell Tissue Kinet.* **21**, 169-182

2. Mikulski, S.M., Viera, A., Ardelt, W., Menduke, H., and Shogen, K. (1990). Tamoxifen and trifluoroperazine (stelazine) potentiate cytostatic/cytotoxic effects of P-30 protein, a novel protein possessing anti-tumor activity. *Cell Tissue Kinet.* **23**, 237-246

3. Mikulski, S.M., Ardelt, W., Shogen, K., Bernstein, E.H., and Menduke, H. (1990). Striking increase of survival of mice bearing M109 Madison carcinoma treated with a novel protein from amphibian embryos. *J. Natl. Cancer Inst.* **182**, 151-153

4. Ardelt, W., Mikulski, S.M., and Shogen, K. (1991). Amino acid sequence of an anti-tumor protein from *Rana pipiens* oocytes and early embryos. *J. Biol. Chem.* **266**, 245-251

5. Feldmann, R.J., Bing, D.H., Potter, M., Mainhart, C., Furie, B., Furie, B.C., and Caporale, L.H. (1985). On the construction of computer models of proteins by the extension of crystallographic structures. *Ann. N.Y. Acad. Sci.* **439**, 12-43

6. Greer, J. (1985). Protein structure and function by comparative model building. *Ann. N.Y. Acad. Sci.* **439**, 44-63

7. Greer, J. (1990). Comparative modeling methods: Application to the family of the mammalian serine proteases. *Prot. Structure Function and Genetics* **7**, 317-334

8. Richards, F.M., and Wyckoff, H.W. (1971). The Enzymes (P.D. Boyer, ed.), 3rd edition, Vol. 4, pp. 647-806, Academic Press, NY

9. Barkakoti, N., Moss, D.S., and Palmer, R.A. (1982). Ribonuclease-A: Least-squares refinement of the structure at 1.45 Å resolution. *Acta Cryst.* **B38**, 2210-2217

10. Wlodawer, A., Bott, R., and Sjölin, L. (1982). The refined crystal structure of ribonuclease A at 2.0 Å resolution. *J. Biol. Chem.* **257**, 1325-1332

11. Wlodawer, A., and Sjölin, L. (1983). Structure of ribonuclease-A: Results of joint neutron and X-ray refinement at 2.0 Å resolution. *Biochemistry* **22**, 2720-2728

12. Wyckoff, H.W., Tsernoglou, D., Hanson, A.W., Knox, J.R., Lee, B., and Richards, F.M. (1970). The three-dimensional structure of ribonuclease-S. *J. Biol. Chem.* **245**, 305-328

13. Powers, T.B. (1976). Ph.D. Thesis, Yale University

14. Greer, J. (1981). Comparative model-building of the mammalian serine proteases. *J. Mol. Biol.* **153**, 1027-1042

15. Read, R., Brayer, G.D., Jurášek, L., and James, M.N.G. (1984). Critical evaluation of comparative model building of *Streptomyces griseus* trypsin. *Biochemistry* **23**, 6570-6575

16. Seqsee 2.0 (unpublished) Wishart, D., Boyko, R., and Sykes, B. (Distributors) Department of Biochemistry, University of Alberta, Edmonton, Alberta

17. Dayhoff, M.O. (1979). Atlas of Protein Sequence and Structure, Vol. 5, Suppl. 3, National Biomedical Research Foundation, Washington, DC

18. McLachlan, A.D. (1972). Repeating sequences and gene duplication in proteins. *J. Mol. Biol.* **64**, 417-437

19. Janin, J., Wodak, S., Levitt, M., and Maigret, B. (1978). Conformation of amino acid side-chains in proteins. *J. Mol. Biol.* **125**, 357-386

20. Moult, J., and James, M.N.G. (1986). An algorithm for determining the conformation of polypeptide segments in proteins by systematic search. *Proteins* 1, 146-163

21. FRODO 2.4. Jones, T.A. *Meth. Enzym.* **115**, 157-171 (Adapted for Silicon Graphics Iris Workstations)

22. Cambillau, C.M. (1987). "Documentation for TOM, Molecular Graphics Program for the Iris", CRMC2, Campus Luminy, Case 913 Marseille, France

23. QUANTA 3.0. Polygen Corporation, 200 Fifth Avenue, Waltham, MA 02254

24. Richards, F.M., and Wyckoff, H.W. (1973). Atlas of Molecular Structures in Biology (D.C. Phillips and F.M. Richards, ed.), Vol. 1, Clarendon Press, Oxford

25. Wodak, S.Y., Liu, M.Y., and Wyckoff, H.W. (1977). The structure of cytidylyl (2',5') adenosine when bound to pancreatic ribonuclease S. *J. Mol. Biol.* **116**, 855-875

26. Pavlovsky, A.G., Borisova, S.N., Borisov, V.V., Antonov, I.V., and Karpeisky, M.Y. (1978). The structure of the complex of ribonuclease-S with fluoride analogue of UpA at 2.5 Å resolution. *FEBS Lett.* **192**, #2, 258-262

27. Borkakoti, N. (1983). The active site of ribonuclease-A from the crystallographic studies of ribonuclease-A-inhibitor complexes. *Eur. J. Biochem.* **132**, 89-94

28. Wlodawer, A., Miller, M., and Sjölin, L. (1983). Active site of RNase: Neutron diffraction study of a complex with uridine vanadate, a transition-state analog. *Proc. Natl. Acad. Sci.* **80**, 3628-3631

29. Usher, D.A., Richardson, D.I., and Ekstein, F. (1970). Absolute stereochemistry of the second step of ribonuclease action. *Nature (London)* **228**, 663-665

30. Usher, D.A., Erenrich, E.S., and Ekstein, F. (1972). Geometry of the first step in the action of ribonuclease-A. *Proc. Natl. Acad. Sci. (USA)* **69**, 115-118

31. Witzel, H., and Barnard, E.A. (1962). Mechanism and binding sites in the ribonuclease reaction II. Kinetic studies on the first step of the reaction. *Biochem. Biophys. Res. Comm.* **7**, #4, 295-299

CHAPTER 3

THE REFINED 1.7 Å X-RAY CRYSTALLOGRAPHIC STRUCTURE OF P-30 PROTEIN, AN AMPHIBIAN RIBONUCLEASE WITH ANTI-TUMOR ACTIVITY†

INTRODUCTION

P-30 Protein (Onconase)‡ is an abundant amphibian ribonuclease (RNase) that is isolated from the oocytes and early blastomeres of *Rana pipiens.* It is a member of the pancreatic RNase superfamily (Ardelt *et al.*, 1991) and it is one of several RNases that display a cytotoxic activity *in vivo* (Mikulski *et al.*, 1990a)‡. P-30 is active against a number of the tumor cell lines tested *in vitro.* At concentrations of 1.5 µg/mL, P-30 can reduce the number of cells that undergo mitosis by up to 90% (Darzynkiewicz *et al.*, 1988; Mikulski *et al.*, 1990b). It is a distant relative of the extensively characterized RNase A and seminal RNase enzymes (Beintema *et al.*, 1988). The seminal RNase homodimer shows a similar growth inhibition (up to 95%) at 50 µg/mL and RNase A is still weaker as a cytotoxin (Tarnowski *et al.*, 1976). The cytotoxic mechanism of P-30 and related enzymes is under investigation. While it is known that RNase A is a potent cytotoxin when injected (Saxena *et al.*, 1991) or when targeted for entry (Ryback *et al.*, 1991) into *Xenopus* oocytes; it is not known if intact, active RNase A enters susceptible cells *in vivo.* The seminal RNase homodimer is internalized by both susceptible and resistant cell lines (Vescia *et al.*, 1980). P-30 binds to 9L glioma cells with two separate affinities (K_d=6.2x10^{-8}M and K_d=2.5x10^{-7}M) and its IC_{50} is similar to the low affinity K_d (Wu *et al.*, 1993). Wu *et al.* (1993) have also shown that P-30's RNase activity and cytotoxicity are inhibited by the alkylation of histidine residues. These results suggest P-30 has an

---

† A version of this chapter has been published. Mosimann SC, Ardelt W. & James MNG. 1994. The refined 1.7 Å x-ray crystallographic structure of p-30 protein, an amphibian ribonuclease with anti-tumor activity. *J. Mol. Biol. 236*, 1141-1153.

‡P-30 is currently undergoing phase II human clinical trials (U.S.A.).

intracellular site of action that depends upon its ribonucleolytic activity. If this is true, it may be significant that amphibian RNases are not inhibited by mammalian placental ribonuclease inhibitor (PRI) and that the mammalian RNases are not inhibited by the amphibian counterpart to PRI (Beintema *et al.*, 1988). This may explain P-30's enhanced cytotoxicity when compared to the mammalian enzymes. Alternatively, it may be P-30's distinctive RNase activity or its specific binding to the surface of susceptible cells that contribute to its increased cytotoxicity.

In this work, the X-ray crystallographic structure solution and refinement of P-30 is reported at 1.7 Å resolution. The molecular conformation is described with an emphasis on the stereochemistry of the active site. The discussion assesses the accuracy of the previously published comparative molecular model (Mosimann *et al.*, 1992).

METHODS

P-30 was isolated and purified from *Rana Pipiens* oocytes and early embryos using previously described protocols (Ardelt *et al.*, 1991). Orthorhombic crystals (space group, $P2_12_12_1$) of P-30 were grown according to Mosimann *et al.* (1992). The unit cell parameters for the high-resolution native data set used in the refinement are $a = 40.53 \pm 0.08$ Å, $b = 69.64 \pm 0.09$ Å, $c = 32.52 \pm 0.08$ Å and $\alpha = \beta = \gamma = 90°$ The calculated solvent content for this crystal is 36.3% and the specific volume, V_m, is 1.94 $Å^3$/Da (Matthews, 1968).

(a) *Structure determination*

The structure of P-30 was determined using the multiple isomorphous replacement (MIR) method. The X-ray intensity data were collected locally on San Diego Multiwire Systems (SDMS) multiwire proportional counters. Graphite monochromated $CuK_{\bar{\alpha}}$ X-rays were generated with a Rigaku Rotoflex RU-200BH rotating anode source operated at 40 kV and 150 mA. The data were processed to 1.7 Å for the native protein and to

maximum resolutions of 2.4 Å, 2.7 Å and 3.0 Å for the three heavy atom derivatives using the SDMS software (Hamlin, 1985; Howard *et al.*, 1985). Statistics for the processing of native and derivative data sets are presented in Table 3.1. The heavy atom binding sites were located from difference Patterson maps and subjected to both positional and occupancy parameter refinement with the program MLPHARE (Otwinowski, 1991). The MIR phasing statistics for each of the heavy atom derivatives are presented in Table 3.2. The absolute hand was determined using the anomalous signal of the trimethyl lead acetate derivative and confirmed by the presence of right-handed helical density in the resulting electron density map. The 2.7 Å MIR electron density map was calculated using figure of merit (FOM) weighted structure factor amplitudes and phases having a mean FOM of 0.68. This map was of sufficient quality to enable the tracing of the complete polypeptide chain. The comparative molecular model of P-30 (Mosimann *et al.*, 1992) was fitted to the electron density on an SGI 4D-25 workstation using a locally modified interactive graphics program TOM Alberta/CalTech 2.8.0 (based on FRODO: Jones, 1985). The three refined heavy atom positions are located on the surface of the polypeptide chain tracing in the MIR map and each is within 5 Å of P-30 and a symmetry related molecule.

During the MIR structure determination of P-30, a number of attempts were made to solve the structure by molecular replacement. Both the comparative molecular model of P-30 (Mosimann *et al.*, 1992) and the crystallographic structure of the phosphate containing bovine RNase A (Wlodawer *et al.*, 1983) were used as search models for rotation function calculations with the following programs: YORROT (Dodson, E.J., 1985), MOLECULAR REPLACEMENT PROGRAM PACKAGE (Alzari & Navaza, 1990) and X-PLOR (Brünger *et al.*, 1987). Additional search models representing truncated versions of the original molecular models were also tested with YORROT. For each search model a series of rotation function maps were calculated as several parameters were varied: (1) the resolution range (the maximum resolution was varied between 5.0Å and 2.0Å; the minimum resolution was varied between 15.0Å and 5.0Å) and

Table 3.1

Summary of data collection statistics for P-30 native and derivative data sets+

	Res. (Å)	No. Ref.	Unique Ref.	% Complete	I/σ(I)	R_m‡	unit cell lengths (Å) *a*	*b*	*c*
Native	1.7	43858	9495	89.4	11.0	0.050	40.53	69.64	32.52
Native	2.1	22458	5104	88.7	20.0	0.029	40.51	69.63	32.53
Pt Deriv.	2.4	13032	3833	98.3	20.7	0.051	40.59	69.61	32.51
Pb Deriv.	2.7	12549	2670	96.9	10.4	0.065	40.48	69.65	32.51
V Deriv.	3.0	8248	1833	89.4	22.9	0.036	40.65	69.73	32.48

+ Space Group $P2_12_12_1$; wavelength ($CuK_{\bar{\alpha}}$) = 1.5418Å

‡ $R_m = \Sigma_{hkl}\Sigma|I_{ave} - I_{obs}|/\Sigma|I_{ave}|$

Pt = platinum ethylenediamine dichloride; Pb = trimethyl lead acetate; V = oxo-vanadium diflouride.

Table 3.2

Multiple Isomorphous Replacement phasing statistics for P-30 at 2.7Å resolution

	sites	R_{diff}[τ]	fractional coordinates			B	Occ.	Phasing
			x	y	z	($Å^2$)	(e^-)	Power[‡]
Pt	1	0.049	0.0480	0.2975	0.4189	44.16	64.87	1.21
Pb	1	0.158	0.2235	0.2141	0.0431	18.77	33.66	1.77
V	1	0.091	0.1208	0.3555	0.2107	31.66	28.50	1.61

Mean Figure of Merit = 0.678

Pt = platinum diethylamine dichloride; Pb = trimethyl lead acetate; V = oxo-vanadium diflouride.

τ $R_{diff} = \Sigma||F_{PH}|-|F_{nat}||/\Sigma|F_{nat}|$

‡ Phasing Power $= \sqrt{\Sigma F_H^2} \; / \; \sqrt{\Sigma(|F_{PH\,obs}|-|F_{PH\,calc}|)^2}$

(2) the threshold intensity of the measured data and (3) the integration radii of the rotation function map. A consistent solution to the rotation function was not obtained using any of these search models. In general, there was a lack of discrimination between the highest peak in the rotation function map and other peaks within the same map.

(b) *Structure refinement*

The initial model of P-30 was subjected to molecular dynamics refinement using the program X-PLOR (Brünger, *et al.*, 1987). Data between 20.0-2.0 Å resolution with intensities greater than $3\sigma|I|$ were selected for the refinement using the simulated annealing protocol described in the X-PLOR manual. The protocol includes: (1) energy minimization using restraints to the X-ray data, (2) a molecular dynamics simulation at 2000 K, (3) a slow cooling to 300 K, (4) positional parameter refinement and (5) an individual B-factor ($B = 8\pi\overline{u^2}$, where $\overline{u^2}$ is the mean-square vibrational amplitude) refinement. During this procedure the crystallographic R-factor ($R = \Sigma||F_o|-|F_c||/\Sigma|F_o|$, where $|F_o|$ and $|F_c|$ are the observed and calculated structure factor amplitudes) was reduced from 0.543 to 0.275. The resulting model coordinates were then subjected to five cycles of restrained least-squares refinement using PROLSQ (Hendrickson & Konnert, 1981) prior to manual refitting. For this and each subsequent refitting session the model was displayed using TOM Alberta/Caltech 2.8.0 on the interactive graphics system with two types of electron density maps: (1) a map calculated with FOM weighted coefficients ($2|F_o|-|F_c|$) and calculated phases and (2) a map computed with FOM weighted difference coefficients ($|F_o|-|F_c|$). In all cases the FOM weighted coefficients were generated with the program SIGMAA (Read, 1986). At several points in the refinement standard omit maps were calculated. In these maps selected atoms were omitted from a short (5-9 cycle) refinement and the subsequent map calculation. The resulting electron density in the region of the omitted atoms represents a relatively unbiased view of the true electron density at these positions. During the initial manual refitting session the amino acids Pyr1 (pyroglutamyl),

Asp2, Thr25, Asn26 and Leu27 occupied regions of ambiguous electron density and were removed from the model. These residues were then rebuilt into the relatively unbiased omit density during the following refitting sessions. After a total of 23 cycles of PROLSQ refinement and three manual refits the R = 0.233 for data between 10.0-2.0 Å having $I \geq 3\sigma|I|$. During subsequent refitting sessions, a sulfate anion was added to the active site and 21 ordered solvent molecules were included. The active site anion was modeled as a sulfate due to its presence in the crystallization medium (and the absence of phosphate). The sulfate oxygens were fit to the electron density and refined throughout (final B-factors range from 24Å^2 to 26Å^2). This reduced the R-factor to 0.209 after a total of 47 cycles of restrained least-squares refinement using 10.0-2.0 Å data with $I \geq 3\sigma|I|$.

The model was further refined against 10.0-1.7 Å data having $I \geq 1.5\sigma|I|$ and additional ordered water molecules were included in the model. Throughout the refinement no effort was made to refine water molecules with alternate positions and/or variable occupancies. The final model contains all 104 amino acids, a sulfate anion and 96 ordered water molecules and has an R-factor of 0.178. The stereochemical parameters associated with the final cycle of least-square refinement are given in Table 3.3. The P-30 coordinates have been deposited in the Protein Data Bank (pdb1onc.ent) at Brookhaven National Laboratory.

RESULTS

(a) *Quality of the refined structure*

In the final $2|F_o|-|F_c|$ map (1s contour level) there is continuous electron density for all main-chain atoms. Two or more side-chain atoms of the residues Lys8, Arg40, Glu42, Lys44, and Lys78 do not have associated electron density. Lys8 and Lys44 have no density beyond C^γ and Lys78 has none beyond C^δ at this contour level. Arg40 and Glu42

Table 3.3

Stereochemical statistics for the final cycle of restrained least-square refinement

R-factor	0.178 for 8653 reflections > 1.5σ|I| between 10.0 - 1.7 Å resolution		
	0.231 for all 9495 unique reflections between ∞ - 1.7 Å resolution		
Model			B-factor (mean)
	826 protein atoms		13.4 Å^2
	96 waters		30.2 Å^2
	1 sulfate anion		23.9 Å^2
Restraints		rmsd(Å)	Target(Å)
	1 - 2 bond distance	0.016	(0.016)
	1 - 3 bond distance	0.030	(0.022)
	1 - 4 planar distance	0.034	(0.030)
	planar groups	0.009	(0.010)
	chiral volumes	0.216	(0.200)
	non-bonded contacts		
	single torsion contacts	0.218	(0.200)
	multiple torsion contacts	0.167	(0.200)

rmsd = $\sqrt{\Sigma(X_x - \bar{x})^2 / m}$; Target is the stereochemical weight assigned during the last cycle of least squares refinement.

have electron density have electron density beyond C^{β} but it is weak and discontinuous. The Arg40-Glu42 residues form intermolecular contacts with Thr25'-Leu27' and the electron density associated with these residues is shown in Figure 3.1.

A Ramachandran plot of the P-30 dihedral angles is presented in Fig. 3.2. P-30 has three glycine residues and 101 non-glycine residues. In the case of the glycine residues, two (Gly50 and Gly102) participate in reverse turns and adopt conformations that are most commonly associated with glycine residues while the third (Gly100) is part of the C-terminal strand β5. Gly50 and Gly102 each occupy the *i+2* position of their respective reverse turns (Table 3.4). Of the remaining 101 residues five (His29, Cys30, Cys75, Asn92, Gln93) have left-handed helical conformations. All of these residues are in reverse turns occupying the second or third positions (Table 3.4). Residues 91 to 94 form a type I' (Sibanda & Thornton, 1985) reverse turn (Fig. 3.3.). The short contact (3.01 Å) between the C^{β} of Gln93 (*i+2*) and the carbonyl oxygen of Asn92 (*i+1*) that is common to this conformation is stabilized by several hydrogen bonding interactions: Glu91 O to Ala94 N (3.21 Å), Gln93 N to Ile51 O (2.69 Å) and Gln93 $O^{\varepsilon 1}$ and Wat119 (2.76 Å). Thr71 has a conformation ($\phi = -95°$, $\psi = -162°$) that lies slightly outside the allowed β-conformation. This residue participates in an Asx type turn in which the side-chain $O^{\varepsilon 1}$ forms a hydrogen bond to the main-chain nitrogen (3.11 Å) of the residue *i+2* in the polypeptide (Arg73 N). The Thr71 N also donates a hydrogen bond to the Lys76 O (2.97 Å).

The B-factor versus sequence number profile of P-30 is shown in Fig. 3.4. In general, the lowest thermal parameters in the structure are associated with the β-strands while the residues with the highest thermal parameters comprise loop regions. Eleven residues of P-30 have average B's greater than 15 $Å^2$. Of these, the only ones that belong to a helix or a β-strand are Arg40 and Glu42 (helix α3). Six of the eleven residues occur in two loops (Ser24-Leu27, Ser72-Arg73) that are covalently linked by the nearby Cys30 to Cys75 disulfide bridge. This results in a cluster of relatively mobile residues found at

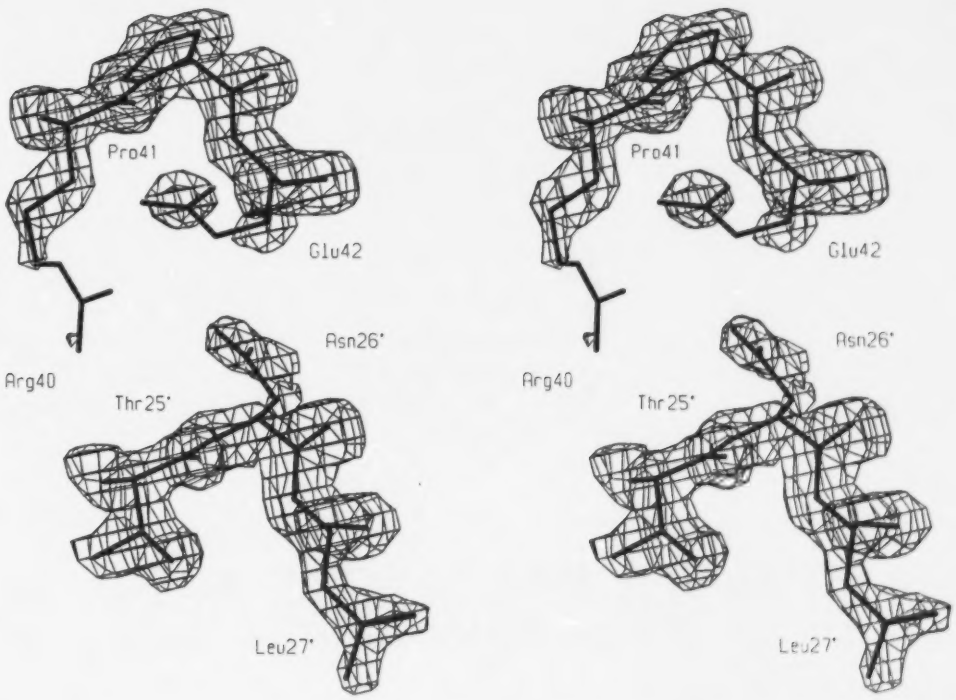


Figure 3.1 The $2|F_o|-|F_c|$ electron density (1σ contour) associated with Arg40-Glu42 and Thr25'-Leu27' of a symmetry related molecule. This represents one of the least well-defined regions of electron density in the P-30 structure.

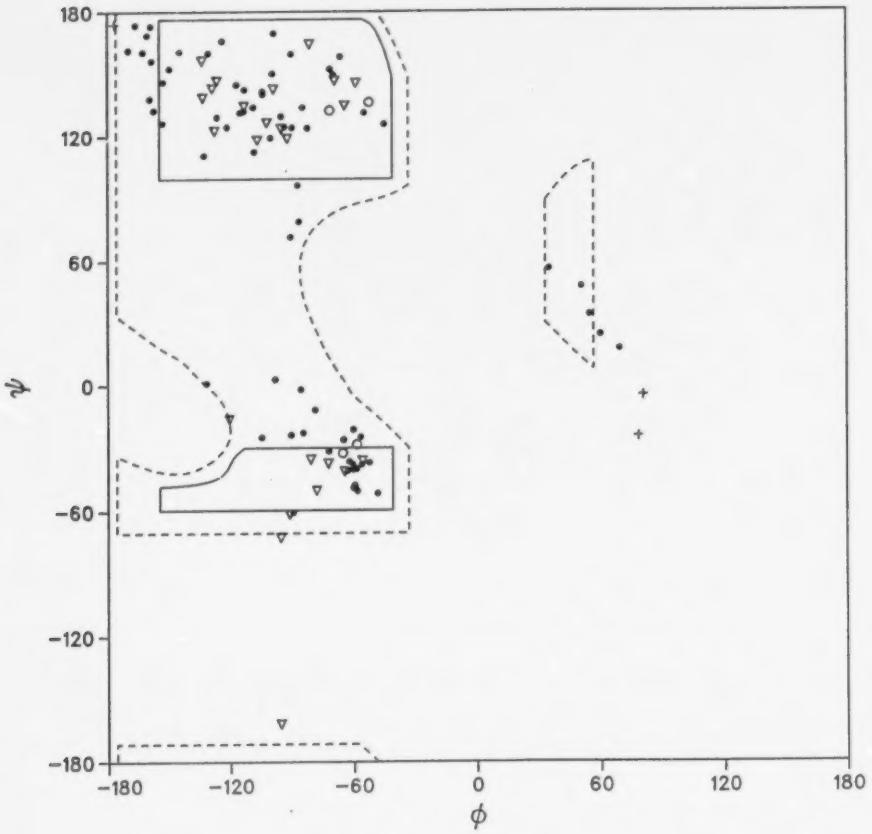


Figure 3.2 The Ramachandran plot of the completed P-30 model is presented with the various allowed regions of conformational space indicated in boxes. Thr71 (ϕ = -95°,ψ = -162°) and Gln93 (ϕ = 69°,ψ = 17°) are the only non-glycine residues that occurs in mildly disallowed regions of the plot. The crosses are glycine residues, the inverted triangles represent amino acids with branched side chains, the open circles correspond to proline residues and closed circles are used for the remaining residues.

Table 3.4

Reverse turns in P-30

Residues	Type	ϕ_2 ψ_2 ϕ_3 ψ_3	H bond distance (Å) $O_i \cdots N_{i+3}$
Phe28-Lys31	I'	35, 56, 55, 30	2.97
Cys48-Ile51	II	-54, 131, 81, -5	3.01
Arg73-Lys76	II	-52, 136, 60, 24	3.07
Glu91-Ala94	I'	51, 48, 69, 17	3.21

The reverse turns have been classified according to Sibanda & Thornton, 1985.

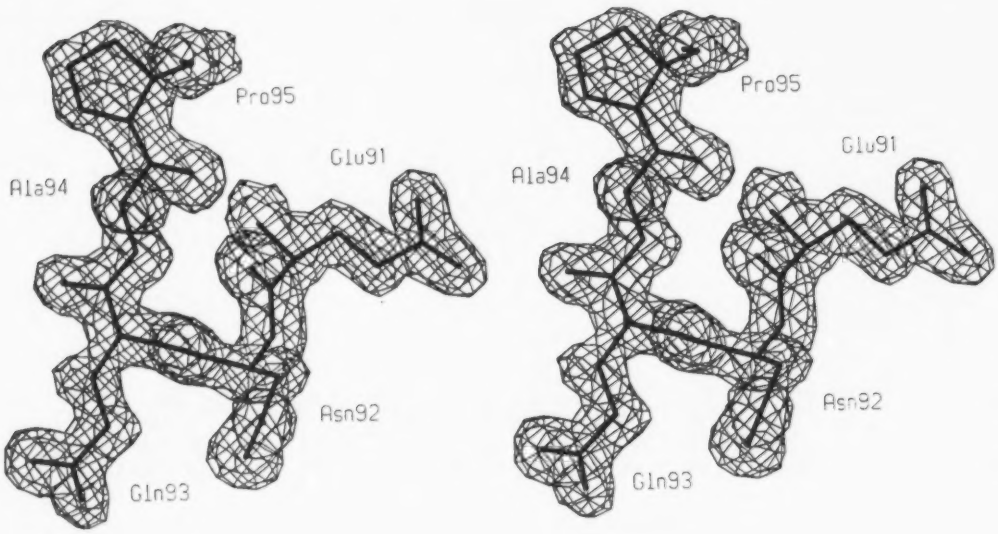


Figure 3.3 The $2|F_o|-|F_c|$ electron density (1σ contour) belonging to Gln93 (ϕ = 69°, ψ = 17°). Gln93 occupies the *i+2* position of this type I' turn, a position typically occupied by glycine. As a result, Asn92 O forms a short intramolecular contact (3.01 Å) with the Gln93 C^{β}.

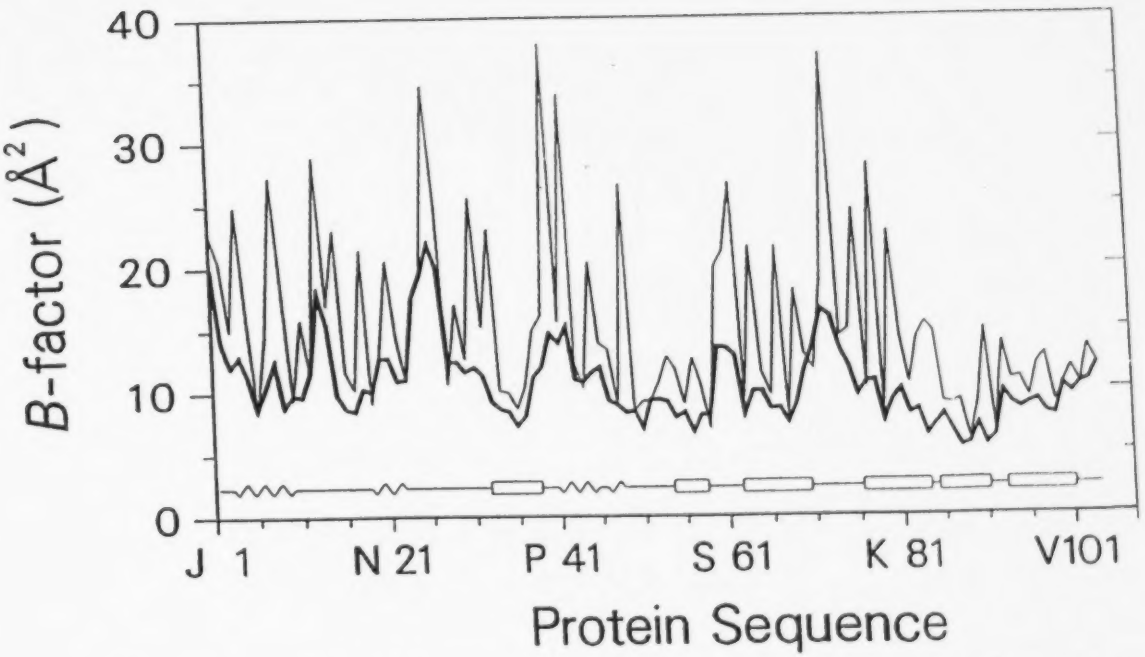


Figure 3.4 A plot of the average thermal parameter (B-factor) against the amino acid sequence number of P-30. The thick line represents the average main-chain B-factor while the thin line represents the average side-chain B-factor: P-30's secondary structures are indicated above the sequence numbers; rectangles represent β-strands, waves represent helices and straight lines indicate turns and loops. The highest average B-factors are associated with several loops (Thr25-Leu27, Arg40-Glu42 and Ser72-Arg73) that are involved in intermolecular contacts.

one end of the molecule (Fig. 3.1). Several of the other residues that have relatively high average main-chain B-factors make intermolecular van der Waal's contacts with these mobile loops. Asn13' contacts Thr71 (3.92 Å) and Glu42' contacts Asn26 (3.76 Å). The final residue with an average main-chain B-factor greater than 15 $Å^2$ is the N-terminal residue, Pyr1. The largest B-factor for a protein atom is 44.7 $Å^2$ (Arg40 $N^{\eta 1}$) and the highest average side-chain B-factors are associated with Asn26, Arg40, Glu42 and Arg73. Other side chains with high average B-factors are Lys8, Lys49, Glu62 and Lys78. In each of these cases, the side chains are relatively long and extend into a disordered solvent channel.

There are 96 full occupancy solvent molecules in the P-30 model. They have B-factors ranging from 12 to 51 $Å^2$ (the highest in the structure). The water molecules are approximately equally distributed within this range. There are 16 waters with B's less than 20 $Å^2$, 28 waters with B's between 20-30 $Å^2$, 35 waters with B's between 30-40 $Å^2$ and 17 waters with B's greater than 40 $Å^2$. The solvent molecules were refined with full occupancy and variable B-factors in all cases.

(b) *Molecular conformation of P-30 Protein*

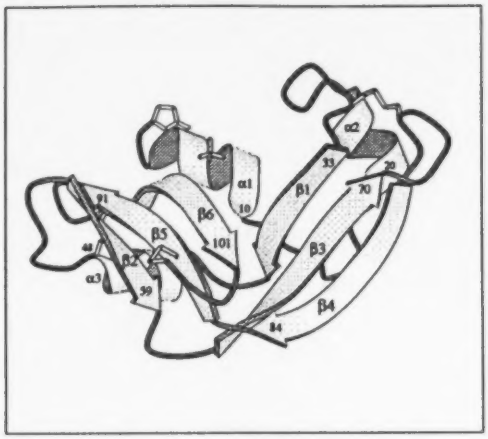
P-30 is a single domain protein that is presently the smallest known member of the pancreatic RNase superfamily. It contains 104 amino acids and has a calculated M_r of 11,834 Da (pH 7.0, His protonated). The fold of its polypeptide chain is shown in Fig. 3.5a and as stick model in Fig. 3.5b. P-30's fold is closely related to that of RNase A. It is a predominantly antiparallel β-structure with a prominent N-terminal α-helix that makes contacts to each of its two lobes.

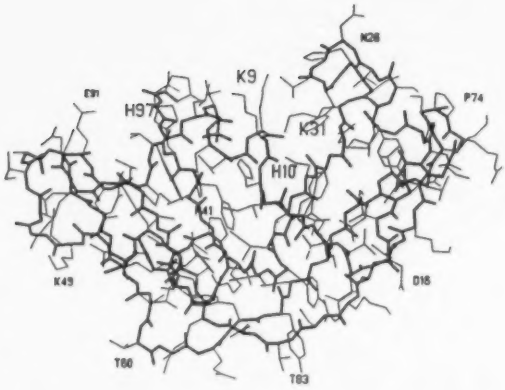
The secondary structure of P-30 has been analyzed with the DSSP program (Kabsch & Saunders, 1983), giving the following proportions: 40.4% antiparallel β-strand, 20.2% helix, 39.4% turns and aperiodic segments (loops). The intramolecular

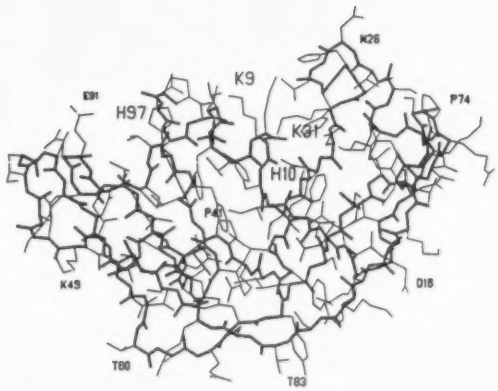
Figure 3.5a A Molscript (Kraulis, 1991) ribbons diagram of the molecular conformation of the P-30 protein. The helices and strands are labeled sequentially as they occur in the primary sequence. Three of the four disulfide bonds are visible while the fourth (Cys19-Cys68) is hidden behind strand β3. The N-terminal pyroglutamyl residue and the active site sulfate anion are shown with all their bonds.

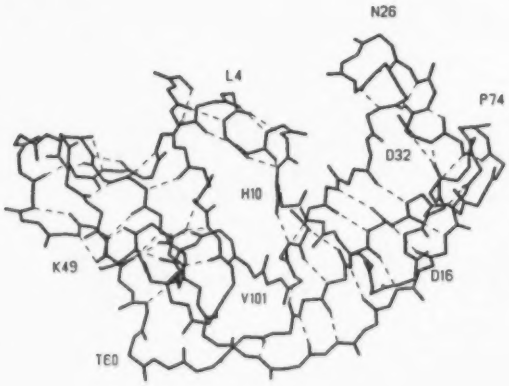
Figure 3.5b An all atom representation of the P-30 molecule. The main-chain atoms are connected by thick lines while the thin lines represent the bonds between side-chain atoms. The sulfate anion at the active site is also shown in thin lines.

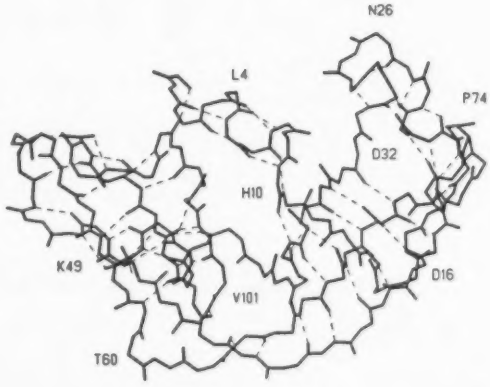
Figure 3.5c The main-chain tracing of the P-30 molecule. The solid thick lines connect all main-chain atoms and the thin dashed lines represent hydrogen bonds between pairs of main-chain atoms (2.50 Å ≤ d ≤ 3.30 Å).


α2
33
20
α1
β1
10
70
β6
91
β5
101
β3
4
β2
α3
39
β4
84


N26
K9
E91
H97
P74
K31
H10
41
K49
D16
T60
T83


N26
K9
E91
H97
P74
K31
H10
K49
D16
T60
T83


N26
L4
P74
D32
H10
K49
D16
V101
T60


N26
L4
P74
D32
H10
K49
D16
V101
T60

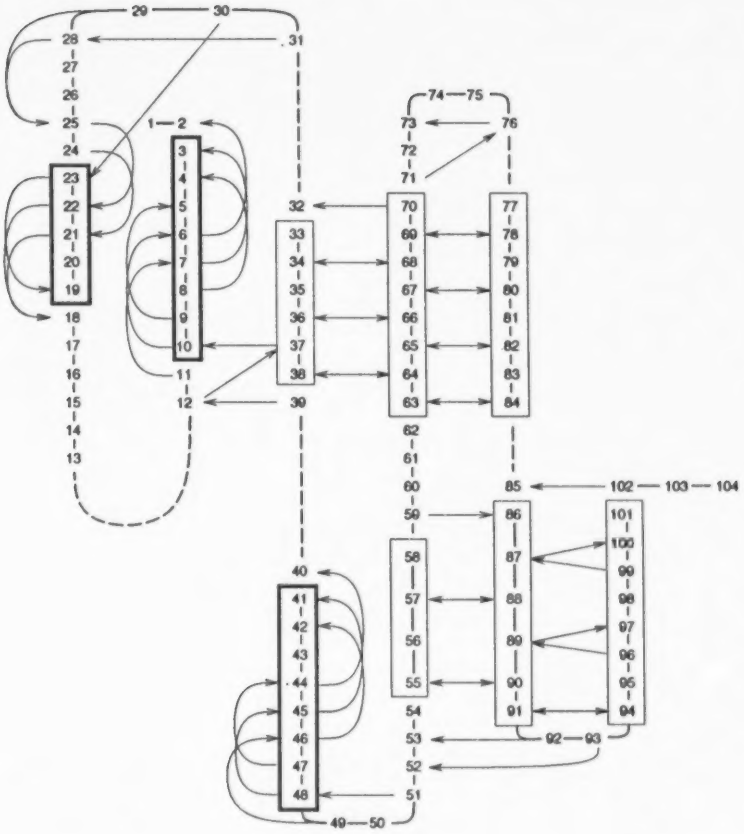


Figure 3.6 The intramolecular main-chain hydrogen bonding network for P-30. The β-strands are indicated by thin boxes and the helices by thick boxes. The arrows represent hydrogen bonds (2.50 Å ≤ d ≤ 3.30 Å) and the arrowheads are directed at the hydrogen bond acceptors. Residues Cys87 and Thr89 of strand β5 each form main-chain hydrogen bonds with two residues from strand β6.

main-chain hydrogen bonding pattern for P-30 is shown in Fig. 3.5c and Fig. 3.6. The antiparallel β-strands are found in two three-stranded β-sheets that give P-30 its bilobate appearance. The first of these sheets contains strand β1 (Lys33-Tyr38), strand β3 (Phe63-Val70) and strand β4 (Tyr77-Asn84). The other β-sheet is several residues smaller and consists of the strand β2 (Lys55-Leu58), β5 (Phe86-Glu91) and β6 (Ala94-Val101). The strands β4 and β5 are separated by a single residue and may alternatively be considered a single, twisted β-strand with a β-bulge at Lys85. The main-chain hydrogen bonds that involve strand β6 are the only hydrogen bonds that deviate from those expected for antiparallel β-sheets. Strand β6 contains bulges at Val96-His97 and Val99-Gly100 that give rise to the hydrogen bonding pattern observed in Fig. 3.6. Those residues assigned a helical conformations form helices: α1 (Trp3-His10), α2 (Cys19-Met23) and α3 (Pro41-Cys48). The helix α1 is completely α-helical in nature while helices α2 and α3 are distorted and contain mixtures of α-helix and 3_{10} helix . As indicated in Fig. 3.6, helix α2 contains one i→i+4 hydrogen bond (Cys19 O→Met23 N) and several i→i+3 hydrogen bonds while helix α3 is α-helical from Pro41 to Lys45 and 3_{10} helix from Ala46 to Cys48. The helices α2 and α3 have similar structural roles in the P-30 fold as each packs against one of the previously mentioned three-stranded antiparallel β-sheets with its helix axis approximately perpendicular to the strand direction.

Those residues that do not adopt helical or β-strand conformations account for the remainder of the P-30 structure. These segments include four reverse turns that have i→i+3 hydrogen bonds and several aperiodic segments. The reverse turns are listed in Table 3.4. Of the segments that do not fit the α-helix, β-strand or reverse turn classification only Ile11-Asp18 is longer than four residues. These residues form an irregular loop connecting helices α1 and α2. The first three residues (Ile11-Asn13) of this loop have an extended β-strand like conformation but only form a single hydrogen bond with the neighboring strand β1 (Thr12 N→Ile37 O, 3.10 Å). The other five residues of this loop are involved in reversing the direction of the polypeptide chain. The equivalent loop

in RNase A contains a subtilisin cleavage site that is not present in P-30.

There are three short connections between the lobes of P-30 that may act as a hinge between the two lobes. The residues at these connections include Ser39-Arg40, Thr59-Glu62 and Lys85. The final segments of the P-30 structure with non-helix and non-strand ϕ, ψ torsion angles are the two N-terminal and the three C-terminal residues. The N-terminal Pyr1 and Asp2 are folded back against the helix α1 and Pyr1 is directed into the active site of the enzyme. The C-terminal residues Gly100-Ser103 form a structure closely related to a reverse turn but lacking the i$\rightarrow$i+3 hydrogen bond. The Gly100 C=O bond vector lies in the plane of the β-sheet (β2, β5, β6) whereas the Ser103 N-H bond vector is perpendicular to the plane of the β-sheet. As a result, this turn-like structure does not have an i$\rightarrow$i+3 hydrogen bond. The C-terminal amino acid, Cys104, forms a disulfide bridge with Cys87.

P-30 is a disulfide-rich protein with eight cysteine residues forming four disulfide bonds (Fig. 3.5a,b). Three of the four disulfide bonds have counterparts in RNase A and these involve the following residues: (1) Cys19 and Cys68, (2) Cys30 and Cys75 and (3) Cys48 and Cys90. The fourth disulfide connects Cys87 and Cys104 and is unique to the amphibian members of the RNase superfamily. The respective -S-S- torsion angles for these disulfide bonds are -79°, -77°, -72° and +99°. The Cys19 to Cys68 bond connects the N-terminus of helix α2 to the central strand (β3) of the antiparallel β-sheet that it packs against. Similarly, the Cys48 to Cys90 disulfide tethers the α3 helix to the central strand (β5) of its associated β-sheet. Whereas these two disulfides fulfill structurally similar roles, the other two disulfides are structurally unrelated as indicated by their ϕ, ψ angles and especially their χ_3 torsion angles. In the Cys30 to Cys75 bridge, the cysteine residues are both involved in reverse turns. Cys30 and Cys75 each occupy the i+2 position of their respective reverse turns and each adopts a left-handed helical conformation as a consequence. The Cys87 to Cys104 disulfide links the C terminus to the central strand (β5) of the β2, β5, β6 antiparallel sheet.

(c) *Intermolecular contacts*

Table 3.5 contains a compilation of the direct intermolecular contacts within the P-30 crystal and Fig. 3.7 is an intermolecular packing diagram. The contacts listed in Table 3.5 have been separated into nine van der Waal's contacts (3.20 <d < 3.90 Å) and 14 hydrogen-bonding interactions. All together, these intermolecular contacts involve 29 of the 104 residues in P-30. All of the nine van der Waal's contacts involve one or more residues that are on loops. The contacts involve or are adjacent to residues in previously discussed sequences possessing relatively high main-chain B-factors. The Pro74 side-chain is involved in four separate close contacts. The shortest intermolecular contact occurs between Tyr64 OH and Arg73 C^{β} (3.22 Å). When examining the intermolecular hydrogen bonds in Table 3.5, two interactions are particularly noteworthy. The first involves the Lys55 N^{ζ} and the sulfate oxygen, O4. In Fig. 3.8, the long Lys55 side-chain adopts an almost fully extended conformation (χ_1 =-175°, χ_2 = -171°, χ_3=-154°, χ_4=168°) that allows it to bridge the solvent channel. The side-chain atoms have well-defined electron density and individual B-factors that are below 15 $Å^2$. Lys55 is the only residue that makes an intermolecular contact with the active site (Fig. 3.8). Its N^{ζ} atom is 2.79 Å from the sulfate O4 and 2.73 Å from the Phe98 O. As a result of this position, Lys55 and the sulfate anion sterically block His97 from adopting χ_1 torsion values near 180°. The second intermolecular interaction of note involves the Glu91 side chain. Glu91 $O^{\varepsilon 1}$ makes two separate hydrogen bonds to the side chains of Asp16 and Lys81. In turn, the Asp16 side chain accepts an intramolecular hydrogen bond from Lys81 N^{ζ}. Given the pH of the crystallization buffer (pH 4.5) and the pK_a's of the carboxyl side-chains (~4.5); the most likely effect is that the Asp16 and Glu91 carboxyls share a single proton in a short hydrogen bond (2.49 Å) and have a single net negative charge. Similar carboxyl-carboxylate interactions have been reported for a variety of other systems

Table 3.5

Summary of the intermolecular close contacts and hydrogen bonds for the refined P-30 model

Symmetry operation	Van der Waal's contacts $d < 3.90$ Å				Hydrogen bonds $d < 3.30$ Å			
x, y, z+1	Tyr64 OH	-	Arg73 C^{β}	3.22	Ser24 O	-	Arg40 $N^{\eta 1}$	2.71
	Pro74 C^{β}	-	Glu62 O	3.36	Ser61 O^{γ}	-	Asp32 $O^{\delta 1}$	2.52
	Phe63 C	-	Pro74 C^{γ}	3.77				
	Tyr64 $C^{\gamma 2}$	-	Pro74 C^{δ}	3.56				
	Thr83 $C^{\gamma 2}$	-	Pro74 C^{δ}	3.84				
	Asn26 $N^{\delta 2}$	-	Arg40 $N^{\eta 2}$	3.20				
x+1/2, -*y*+1/2, -*z*	Asp20 C^{β}	-	Val101 O	3.38	Asn21 $O^{\delta 1}$	-	Lys80 N^{ζ}	2.58
	Asp20 $O^{\delta 1}$	-	Ser103 C^{β}	3.30	Asn21 $N^{\delta 2}$	-	Asp67 $O^{\delta 1}$	2.88
	Asn21 C^{γ}	-	Val101 $C^{\gamma 1}$	3.87	Arg73 $N^{\eta 1}$	-	Ser103 O^{γ}	3.05
	Thr25 $C^{\gamma 2}$	-	Lys80 C^{ε}	3.39	Arg73 $N^{\eta 2}$	-	Ser103 O^{γ}	3.05
					Lys76 N^{ζ}	-	Gly102 O	2.58
x+1/2, -*y*+1/2, -*z*+1					Tyr38 OH	-	Lys85 N^{ζ}	2.97
-*x*+1/2, -*y*, *z*+1/2					Lys55 N^{ζ}	-	Phe98 O	2.79
					Lys55 N^{ζ}	-	Sulfate O4	2.75
-*x*+1, *y*+1/2, -*z*+1/2					Asp16 $O^{\delta 1}$	-	Glu91 $O^{\varepsilon 1}$	2.49
					Leu79 O	-	Asn92 $N^{\delta 2}$	2.94
					Lys81 N^{ζ}	-	Glu91 $O^{\varepsilon 1}$	3.29

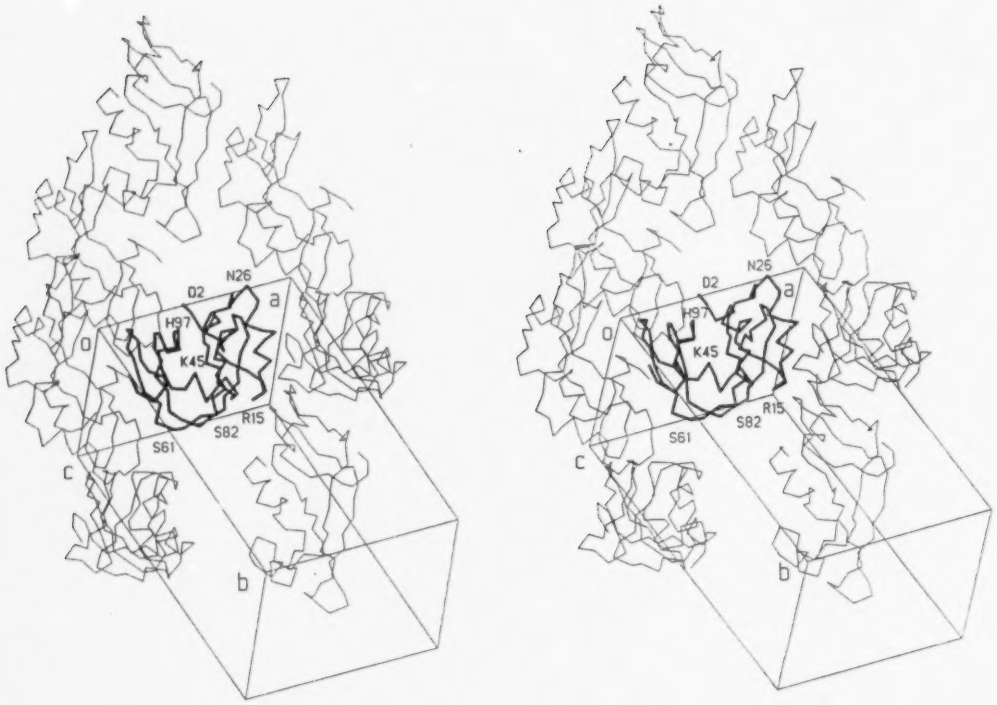


Figure 3.7 The intermolecular packing diagram of orthorhombic P-30. The unit cell is shown as a rectangular box and the thick C^{α} tracing represents a single P-30 molecule. Seven other neighboring P-30 molecules are included in thin lines.

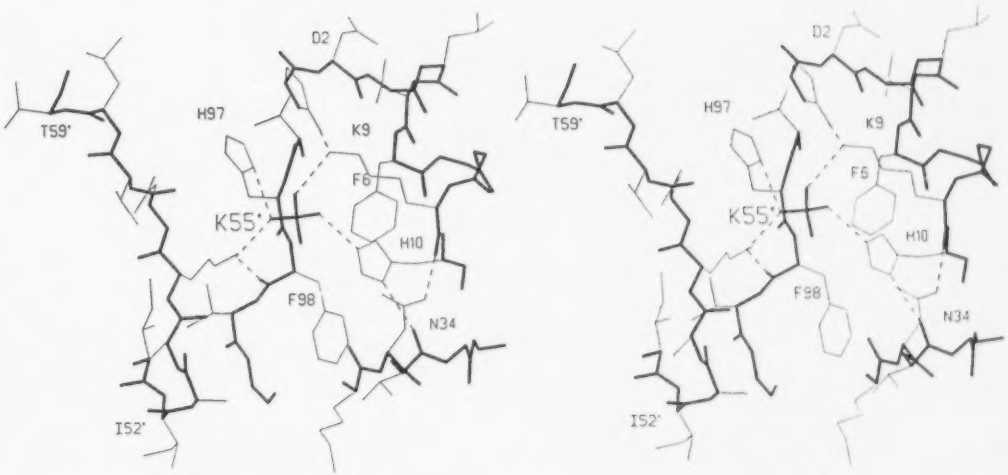


Figure 3.8 The intermolecular contact at the active site is shown as a stick model. The symmetry related Lys55' residue is observed to extend across a solvent channel and into the active site where it makes contacts with the sulfate anion and Phe98. This intermolecular contact prevents His97 from adopting χ_1 torsion angle near 180°. It should be noted that residues adjacent to Lys55' are not involved in intermolecular contacts.

(Sawyer & James, 1982) and very short hydrogen bonds are also found in protein structures (Fraser *et al.*, 1992).

(d) *Active site of P-30 Protein*

The active site is a long cleft that runs approximately parallel to strand β6. It is formed at the junction of helix α1, strand β1 and strand β6. The active site residues are Pyr1, Lys9, His10, Lys31, Thr35, His97 and Phe98. The residues and their associated electron density are shown in Fig. 3.9. The aromatic Phe98, the catalytic His10 and the polar Thr35 side chains line the interior of the cleft. His10 packs against Phe6 (not shown) and the C^{β} of Phe98 and Thr35 packs against the relatively exposed aromatic ring of Phe98. The hydrogen bonding requirements of the His10 and Thr35 side-chains are satisfied by a well-ordered sulfate anion and several ordered water molecules that are non-covalently bound to the active site. The sulfate anion is presumed to be a mimic of the phosphodiester moiety of RNA substrates.

Lys9, His10, His97 and Phe98 form salt bridges and hydrogen bonds with the oxygen atoms of the sulfate anion. Lys9 N^{ζ} is 3.15 Å from the sulfate oxygen, O2 and forms a 3.00 Å hydrogen bond with Pyr1 $O^{\varepsilon 1}$. His10 $N^{\varepsilon 2}$ is directed between the O1 (2.64 Å) and O3 (3.32 Å) atoms of the anion. His97 $N^{\delta 1}$ (2.95 Å) forms a salt bridge and Phe98 N (3.20 Å) form hydrogen-bonded contacts with the sulfate oxygen, O4. This network of salt bridges and hydrogen bonds at the P-30 active site is shown in Fig. 3.10. Active site residues that do not contact the sulfate anion directly include Pyr1, Lys31 and Thr35. The pyroglutamyl residue is folded back against the N-terminal helix and its main-chain nitrogen atom forms a hydrogen bond with Val96 O (2.69 Å). This directs the $O^{\varepsilon 1}$ atom of Pyr1 into the active site where it accepts hydrogen bonds from Lys9 N^{ζ} and Wat306 (2.64 Å). In a model of an enzyme-substrate complex, the Pyr1 side chain is expected to pack against the ribose moiety 5' to the scissile phosphodiester bond. The inclusion of Lys31 as an active site residue is based upon its conserved character within the

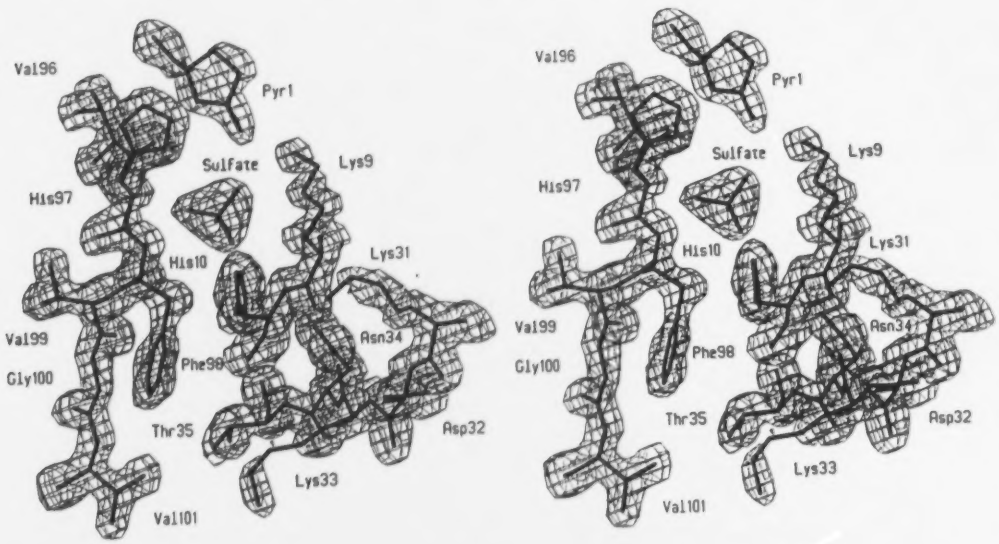


Figure 3.9 The electron density associated with the active site is shown. The contour level is 1σ. A sulfate anion occupies the active site in this orthorhombic crystal form. The refined model is represented by the thick lines. Residues Pyr1, Lys9-His10, Lys31-Thr35 and Val96-Val101 are shown.

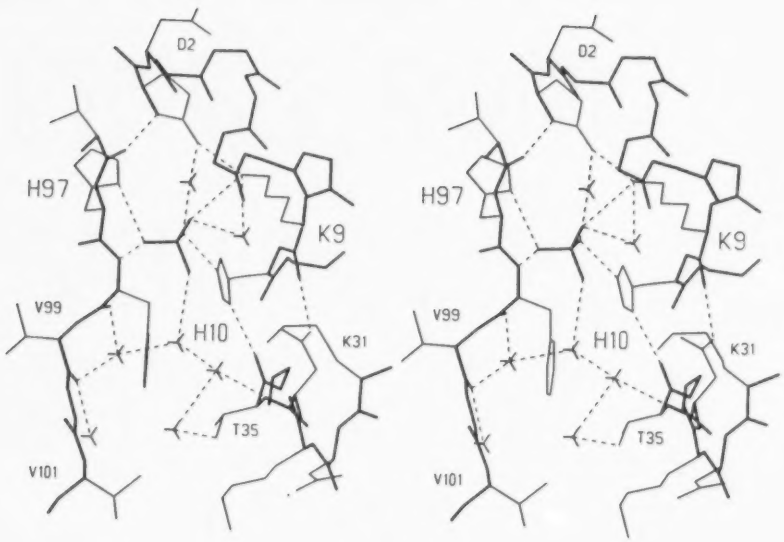


Figure 3.10 An all atom representation of the active site of P-30 is shown with the detailed hydrogen bonding network indicated as dashed lines. The thin lines correspond to side chains while the thick lines represent the bonds between the main-chain atoms of the molecule. The sulfate anion and seven ordered solvent molecules are also included.

pancreatic RNase superfamily and upon experiments in the related RNase A system. Chemical modification studies (Hirs *et al.*, 1965; Raetz & Auld, 1972) and X-ray crystallographic structures of RNase A and RNase S-inhibitor complexes (Richards & Wyckoff, 1973; Wodak *et al.*, 1977; Borkakoti, 1983; Wlodawer *et al.*, 1983) demonstrate that the equivalent residue in RNase A (Lys 41) has a role in catalysis or in substrate binding. In our structure of P-30, Lys31 N^{ζ} is 5.49 Å from the nearest sulfate oxygen (O3). If a similar role is expected for this residue, a conformational change is necessary. The Thr35 N and $O^{\gamma 1}$ atoms form hydrogen bonds with the ordered water molecules Wat117 (3.07 Å) and Wat217 (2.81 Å) respectively within the active site. In the previously mentioned RNase A-inhibitor complexes, the equivalent Thr atoms give rise to the pyrimidine specificity of RNase A by forming hydrogen bonds to O2 and N3 of the base, respectively. P-30 has a similar pyrimidine specificity and it is active against uridine-3',5'-guanylate (UpG) and weakly active against cytidine-3',5'-guanylate (CpG) (Ardelt, unpublished). In contrast, RNase A cleaves all diribonucleotide combinations that begin with a pyrimidine (UpX, CpX where X=A,C,G,U). Three water molecules are involved in the remaining hydrogen bonds to the sulfate anion. Wat306 is 2.47 Å from the sulfate oxygen O1, Wat405 is 3.02 Å from the sulfate oxygen O3 and Wat426 is 2.77 Å from sulfate oxygen O2.

P-30 and RNase A share a two-step, pyrimidine specific endoribonucleolytic activity that is catalyzed by two invariant histidines and an invariant lysine (P-30: His10, Lys31, His97). In RNase A, the first step of the reaction is an intramolecular transesterification that produces a cyclic-2',3'-phosphate intermediate and releases the 3' end of the RNA substrate. The second step is the hydrolytic cleavage of the intermediate to a polyribonucleotide that ends in a 3'-phosphoester of cytidine (Cp) or uridine (Up) (for review see Richards and Wyckoff, 1973). Each of the reaction steps proceeds with an in line geometry and involves a pentacovalent phosphorus transition state. In P-30, His10 and His97 are the equivalent general acid/general base catalysts while Lys9 and Phe98

stabilize the active site anion. In the RNase A-transition state analog X-ray structure (Wlodawer *et al.*, 1983) the invariant lysine interacts with the O2' of the phosphate mimic. The equivalent lysine in P-30 is Lys31 and it is 5.49 Å from the nearest sulfate oxygen. As mentioned above, in order for Lys31 to interact with the P-30 transition state during catalysis, it will have to move.

DISCUSSION

P-30 Protein is a distant relative of bovine RNase A. It is twenty residues smaller and one can align the sequences and achieve 28% identity. Prior to the X-ray structure solution, a comparative molecular model of P-30 was constructed using bovine RNase A coordinates and the aligned sequences (Mosimann *et al.*, 1992). Since comparative molecular model building is dependent upon accurate sequence alignments (Greer, 1985), regions with high sequence identity are expected to show the closest agreement in a comparison of the model and the X-ray structure (Chothia & Lesk, 1986). Fig. 3.11 shows a superposition of the C^{α} atoms of the P-30 model as derived above with the X-ray structure of P-30. The root mean square (RMSD) deviation for the 81 C^{α} atoms that can be superposed within 3 Å is 1.26 Å.

The comparative model of P-30 accurately predicts α-helical and β-strand conformations with two exceptions. The exceptions involve Pyr1 and helix α2. In the model, Pyr1 is part of the N-terminal α-helix. In the X-ray structure Pyr1 folds back against the N-terminal helix becoming part of the active site. This is a significant difference in the modeling of P-30's active site based on RNase A. Helix α2 is located far from the active site. In the X-ray structure, helix α2 is four residues shorter than predicted in the model and it is found systematically shifted by more than 2.0 Å from its predicted position. This difference is related to the conformational differences in the adjacent loop structures and possibly to P-30's intermolecular packing.

The P-30 model does not accurately predict the conformation of loops that possess

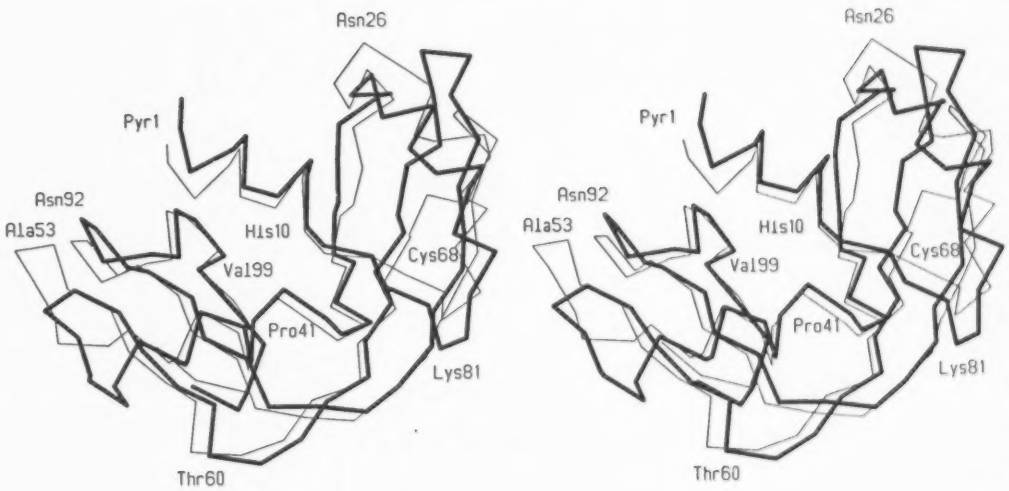


Figure 3.11 The C^{α} superposition of the comparative molecular model (thin lines) and P-30 X-ray crystallographic structure (thick lines). The RMS deviation is 1.26 Å for the 81 C^{α} atoms pairs that superpose within 3 Å.

low sequence identity and are the sites of deletions relative to RNase A. Not surprisingly, these represent the areas of greatest divergence in the X-ray structures of P-30 and RNase A (Fig. 3.11). The loops following helix α3 (P-30: Lys49-Ser54; RNase A: Ser59-Thr70) and strand β3 (P-30: Thr71-Cys75; RNase A: Thr87-Cys95) deserve comment.

In the case of the loop following helix α3, several sequence alignment protocols (Dayhoff, 1979; McLachlan, 1972) position a five residue deletion in P-30 at RNase A residues Thr70-Gln74 (Mosimann *et al.*, 1992). These residues form part of strand β2 in the structure of RNase A. Since the deletion of strand β2 seemed unreasonable, the automated alignment was revised in the comparative molecular model and the deletion was positioned in a loop region of RNase A (Cys65-Asn69). The loop's predicted conformation in P-30 is similar to the RNase A residues Ser59-Ala64. While the alignment revision is justified by the presence of strand β2 in the X-ray structure of P-30, the loop's conformation is radically different in the predicted and real structures. The pairwise positional differences reach 11 Å when comparing the predicted and real structures. Fig. 3.12 is a superposition of the loop region between helix α3 and strand β2 in the real and predicted structures. Such large differences in predicted and real loop conformations have been observed for eucaryotic and procaryotic serine proteinase (Delbaere *et al.*, 1979; Read *et al.*, 1984).

The loop following strand β3 is the site of a three residue deletion in P-30. In the comparative molecular model this deletion was placed between Thr71-Ser72 (Fig. 3.11). As a result, Ser72, Arg73 and Pro74 were aligned with the RNase A residues Lys91, Tyr92 and Pro93. In light of the P-30 X-ray structure, this deletion is now placed between Pro74-Cys75. In the modified alignment Ser72, Arg73 and Pro74 are equivalent to Gly88, Ser89 and Ser90 of RNase A. In this case, the original alignment was chosen over the alternative after examining several related sequences. In the closely related amphibian lectin (53% amino acid identity with P-30), there is no deletion in this loop. The additional residues occur between the P-30 residues Thr71-Ser72; the site of the original deletion. In

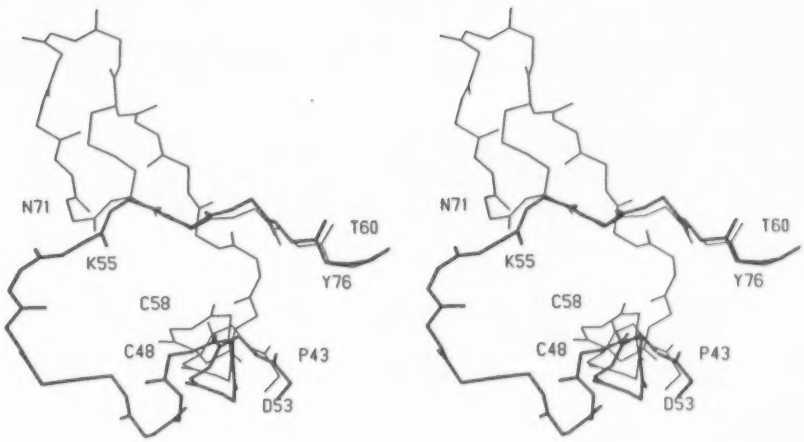


Figure 3.12 The predicted and real conformations of the P-30 loop containing residues Lys49-Asn54. The predicted conformation of the loop is shown in thin lines and the conformation observed in the X-ray structure is shown in thick lines. The region between Cys48 and Lys55 exhibits differences up to 11 Å between equivalent C^{α} atoms.

bovine angiogenin, the sequence equivalent to Lys91, Tyr92 and Pro93 (RNase A) is Ser, Arg, Pro; the same sequence found in P-30. In spite of this evidence for the correct alignment of the deletion, the X-ray structure clearly positions the deletion between residues Pro74-Cys75. The differences in the predicted and real structures of these two loops point to two 'facts of life' in comparative molecular modeling. First, the correct sequence alignment does not guarantee an accurate structure prediction and second, the correct structural alignment is not necessarily the one with the greatest sequence identity.

Finally, it should be pointed out that the Ser72-Pro74 sequence of ONC is involved in several close Van der Waal's contacts with a symmetry related molecule (Table 3.5). Interestingly, these contacts would be relieved by the alignment used in the comparative molecular modeling.

REFERENCES

Alzari, P.M. & Navaza, J. (1990). On the use of the fast rotation function. pp. 303-317. Crystallographic Computing School, Bischenberg.

Ardelt, W., Mikulski, S.M. & Shogen, K. (1991). Amino acid sequence of an anti-tumor protein from Rana Pipiens oocytes and early embryos. *J. Biol. Chem.* **266**, 245-251.

Bernstein, F.C., Koetzle, T.F., Williams, G.J.B., Meyer Jr., E.F., Brice, M.D., Rodgers, J.R., Kennard, O., Shimanouchi, T. and Tasumi, M. (1977). The Protein Data Bank: A computer-based archival file for macromolecular structures. *J. Mol. Biol.*. **112**, 535-542.

Beintema, J.J., Schuller, C., Irie, M. & Carsana, A. (1988). Molecular evolution of the ribonuclease superfamily. *Prog. Biophy. Molec. Biol.* **51**, 165-192.

Borkakoti, N. (1983). The active site of ribonuclease Å from the crystallographic studies of ribonuclease A-inhibitor complexes. *Eur. J. Biochem.* **132**, 89-94.

Brünger, A., Kuriyan, J. & Karplus, M. (1987). Crystallographic R factor refinement by molecular dynamics. *Science* **235**, 458-460.

Chothia, C. & Lesk, A.M. (1986). The relation between the divergence of sequence and structure in proteins. *EMBO J.* **5**, 823-826.

Darzynkiewicz, Z., Carter, S.P., Mikulski, S.M., Ardelt, W. & Shogen, K. (1988). Cytostatic and cytotoxic effects of Pannon (P-30 protein), a novel anticancer agent. *Cell Tissue Kinet.* **21**, 169-182.

Dayhoff, M.O. (1979). Atlas of protein sequence and structure, vol. 5, suppl. 3, Washington, D.C., National Biomedical Research Foundation.

Delbaere, L.T.J., Brayer, G.D. & James, M.N.G. (1979). Comparison of the predicted model of α-lytic protease with the X-ray structure. *Nature* **279**, 165-168.

Dodson, E.J. (1985). Molecular replacement: the method and its problems. Daresbury Study Weekend: Molecular replacement, pp. 33-44.

Greer, J. (1985). Protein structure and function by comparative model building. *Ann. N.Y. Acad. Sci.* **439**, 44-63.

Fraser, M.E., Strynadka, N.C.J., Bartlett, P.A., Hanson, J.E. & James, M.N.G. (1992). Crystallographic analysis of transition-state mimics bound to penicillopepsin: phosphorus-containing peptide analogues. *Biochemistry* **31**,5201-5214.

Hamlin, R. (1985). Multiwire area X-ray diffractometers. *Methods in Enzymol.* **114**, 416-452.

Hendrickson, W.A. & Konnert, J.H. (1981). Stereochemically restrained least squares refinement of macromolecular structure. Biomolecular structure, function, conformation and evolution (Srinivasan, R., ed.), vol. 1, pp. 43-57, Pergamon Press, Oxford.

Hirs, C.H., Holmann, M. & Kyria, J.H. (1965). Dinitrophenylation and inactivation of bovine ribonuclease A. *Arch. Biochem.* **111**, 209-222.

Howard, J., Nielson, C. & Xuong, Ng-H. (1985). Software for a diffractometer with multiwire area detector. *Meth. Enzym.* **114**, 452-472.

Jones, T.A. (1985). Interactive computer graphics:FRODO. *Meth. Enzym.* **115**, 157-171.

Kabsch, W. & Saunders, C. (1983). Dictionary of protein secondary structure: pattern recognition of hydrogen bonded and geometrical features. *Biopolymers* **22**, 2577-2637.

Kraulis, P.J. (1991). Molscript: a program to produce both detailed and schematic plots of protein structures. *J. Appl. Cryst.* **24**, 946-950.

Matthews, B.W. (1968). Solvent content of protein crystals. *J. Mol. Biol.* **33**, 491-497.

McLachlan, A.D. (1972). Repeating sequences and gene duplication in proteins. *J. Mol. Biol.* **64**, 417-437.

Mikulski, S.M., Ardelt, W., Shogen, K., Bernstein, E.H. & Menduke, H. (1990a). Striking increase of survival of mice bearing M109 Madison carcinoma treated with a novel protein from amphibian embryos. *J. Natl. Cancer Inst.* **182**, 151-153.

Mikulski, S.M., Viera, A., Ardelt, W., Menduke, H. & Shogen, K. (1990b). Tamoxifen and trifluoroperazine (stelazine) potentiate cytostatic/cytotoxic effects of P-30 protein, a novel protein possessing anti-tumor activity. *Cell Tissue Kinet.* **23**, 237-246.

Mosimann, S.C., Johns, K.L., Ardelt, W., Mikulski, S.M., Shogen, K. & James, M.N.G. (1992). Comparative molecular modeling and crystallization of P-30 protein: a novel anti-tumor protein of *Rana pipiens* oocytes and early embryos. *Prot. Structure Function and Genetics* **14**, 392-400.

Otwinowski, Z. (1991). Maximum likelihood refinement of Heavy atom parameters. Daresbury CCP4 Study Weekend: Isomorphous replacement and anomolous scattering.

Raetz, C.R. & Auld, D.S. (1972). Schiff bases of pyridoxal phosphate with active center lysines of ribonuclease A. *Biochemistry* **11**, 2229-2236.

Read, R.J. (1986). Improved fourier coefficients for maps using phases from partial structures with errors. *Acta Cryst.* **A42**, 140-149.

Read, R.J., Brayer, G.D., Jurasek L. & James, M.N.G. (1984). Critical evaluation of comparative model-building of *Streptomyces griseus* trypsin. *Biochemistry* **23**, 6570-6575

Richards, F.M. & Wykoff, H.W. (1973). Atlas of molecular structures in biology (Phillips, D.C. & Richards, F.M., eds.). vol. 1, RNase A, Clarendon Press, Oxford.

Ryback, S.M., Saxena, S.S., Ackerman, E.J. & Youle, R.J. (1991). Cytotoxic potential of ribonuclease and ribonuclease hybrid proteins. *J. Biol. Chem.* **266**, 21202-21207.

Sawyer, L. & James, M.N.G. (1982). Carboxyl-carboxylate interactions in proteins. *Nature* **295**, 79-80.

Saxena, S., Ryback, S , Winkler, G., Meade, H., McGray, P., Youle, R. & Ackerman, E. (1991). Comparison of RNases and toxins upon injection into *Xenopus oocytes. J. Biol. Chem.* **266**, 21208-21214.

Sibanda, B. & Thornton, J.M. (1985). β-Hairpin families in globular proteins. *Nature*

316, 170-174.

Tarnowski, G.S., Kassel, R.L., Mountain, I.M., Blackburn, P., Wilson, G. & Wang, D. (1976). Comparison of antitumor activities of pancreatic ribonuclease and its cross-linked dimer. *Cancer Res.* **36**, 4074-4078.

Vescia, S., Tramontano, D., Augusti-Tocco, G. & D'Alessio, G. (1980). *In vitro* studies on selective inhibition of tumor cell growth by seminal ribonuclease. *Cancer Res.* **40**, 3740-3744.

Wlodawer, A. & Sjolin, L. (1983). Structure of ribonuclease A. Results of joint neutron and X-ray refinement at 2.0 Å resolution. *Biochemistry* **22**, 2720-2727.

Wlodawer, A., Miller, M. & Sjolin, L. (1983). Active site of RNase: neutron diffraction study of a complex with uridine vanadate, a transition state analogue. *Proc. Natl. Acad. Sci.* **80**, 3628-3631.

Wodak, S.Y., Liu, M.Y. & Wykoff, H.W. (1977). The structure of cytidylyl -(2,5)-adenosine when bound to pancreatic RNase S. *J. Mol. Biol.* **116**, 855-875 .

Wu, Y., Mikulski, S.M., Ardelt, W., Ryback, S.M. and Youle, R.J. (1993). A cytotoxic ribonucleases: Study of the mechanism of Onconase cytotoxicity. *J. Biol. Chem.* **268**, 10686-10693.

CHAPTER 4

REFINED 2.2 Å X-RAY CRYSTALLOGRAPHIC STRUCTURE OF RECOMBINANT ONCONASE AND COMPARISON WITH ONCONASE AND RNASE A†

INTRODUCTION

Onconase™(ONC) is a small (104 amino acids), basic ribonuclease (RNase) isolated from the eggs of *rana pipiens* (Ardelt *et al.*, 1991). Its amino-acid sequence is distantly related to members of the pyrimidine specific RNase superfamily of vertebrates (Ardelt *et al.*, 1991). A comparative molecular model constructed on the basis of this sequence similarity confirmed that the sequence of ONC makes it a member of the RNase superfamily (Mosimann *et al.*, 1992). ONC displays a specific cytotoxic activity against several tumor cell lines *in vitro* (Darzynkiewicz, Z. *et al.*, 1988) and *in vivo* (Mikulski *et al.*, 1990; Mikulski *et al.*, 1993). Its RNase activity is required for its cytotoxic action (Wu *et al.*, 1993). When compared to RNase A, ONC has a greatly reduced RNase activity (0.001-1.9%) towards various diribonucleotide substrates and exhibits a different substrate specificity (W. Ardelt, unpublished results).

The X-ray crystallographic structure of ONC has been determined at 1.7Å resolution (Mosimann *et al.*, 1994). Its overall fold is similar to RNase A (Zegers *et al.*, 1994), RNase S (Richards & Wykoff, 1971) and angiogenin (Acharya *et al.*, 1994). One novel feature of the ONC structure is the hydrogen bond between the N-terminal pyroglutamyl 1 (Pyr1; 5-oxo-2-pyrollidine carboxylate) $O^{\varepsilon 1}$ atom and Lys9N$^{\zeta}$ at the enzyme's active site.

Recently, recombinant Onconase (rONC) having the N-terminal sequence of ^{f}Met.$_1$(formyl methionine)Glu$_1$Asp$_2$··· has been produced in a bacterial expression system and the expressed enzyme was shown to have no significant ribonucleolytic activity or cytotoxic activity (less than 2% in high-salt conditions, Y. Wu & R.J. Youle, unpublished

results). The rONC X-ray crystallographic structure has been determined and refined at 2.2Å resolution to further investigate the structural basis for its lack of activity. The resulting structure has been compared to that of wild-type ONC. To further our understanding of the ONC system's ribonucleolytic activity and specificity, ONC-deoxyuridine-3',5'-deoxyguanylate (d(UpG)) and rONC-d(UpG) models have been constructed based upon the RNase A-deoxycytidine-3',5'-deoxyadenylate (d(CpA)) crystallographic structure (Zegers *et al.*, 1994).

RESULTS

Quality of the refined structure

The refined rONC structure includes 828 non-hydrogen protein atoms, a single sulfate anion bound in the active site and 72 ordered solvent molecules. The N-terminal $^{f}Met_{-1}$ residue is not ordered in the electron density maps and its position not included in the refined model. Table 4.1 presents the conventional R-factor, the root mean-squared deviation of various stereochemical parameters and the average B-factors for several classes of atoms for the rONC structure. The rONC coordinates will be deposited in the Protein Data Bank at the Brookhaven National Laboratory (Bernstein *et al.*, 1977).

In the final $3|F_o|-2|F_c|$ ($|F_o|$ and $|F_c|$ are the measured and calculated structure factors respectively) electron density map contoured at 1σ, the electron density is continuous and well defined for all main chain atoms excepting Glu1N. At the same contour level, there is no electron density for the Glu1 side-chain and several other solvent exposed side-chain groups have discontinuous electron density. At the 0.8σ contour level the Glu1 side-chain electron density is weak and discontinuous. The electron density associated with three of the twelve lysine side chains (33, 49 and 80) is absent beyond C^{γ}. These residues represent the most-poorly defined regions of electron density in the structure. Figure 4.1 shows a stereo diagram of the electron density associated with residues Glu1 to His10

Table 4.1

Stereochemical statistics for the final cycle of restrained parameter least-square refinement

R-factor	0.165 for 4968 reflections $\|F_o\| > 1.0\sigma\|F_o\|$ between 20.0 - 2.2 Å resolution		
Model			B-factor (mean)
	417 mainchain atoms[1]		15.0 Å^2
	410 sidechain atoms		19.0 Å^2
	72 waters		32.7 Å^2
	1 sulfate anion		26.4 Å^2
Stereochemistry		rmsd[2]	Restraint[3]
	Bond length	0.013Å	0.700
	Bond Angles	1.62°	0.700
	Pseudorotation Angles	13.95°	1.00
	Trigonal planar groups	0.009Å	1.50
	General planar groups	0.008Å	3.00
	non-bonded contacts	0.035Å	5.00
	Temperature factor	2.17Å^2	0.200

[1]The Met -1 residue was not observed in this work and has not been included in the model.
[2]rmsd = $((\Sigma_i(X_i-X_j)^2/N)^{0.5}$, where X_i and X_j are the observed and ideal values and N is the number of values being considered.
[3]Restraint is the relative weight assigned to the stereochemical parameter during the final cycle of refinement.

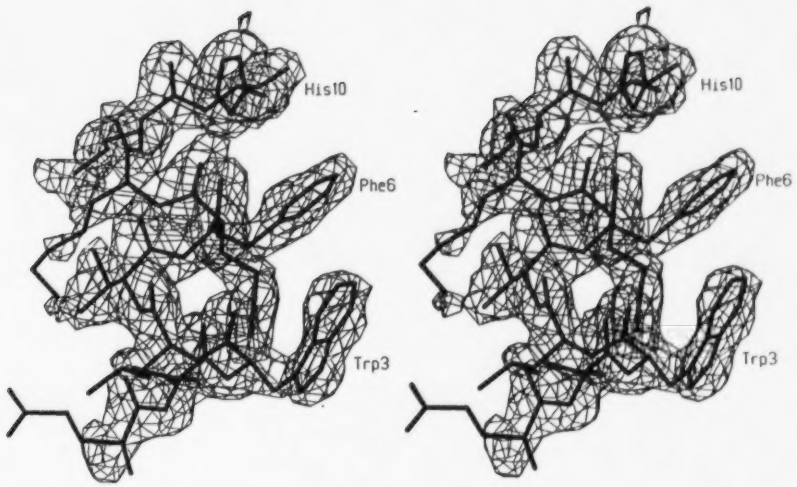


Figure 4.1 The $3|F_o|-2|F_c|\ \alpha_{calc}$ electron density distribution associated with the N-terminal α-helix of rONC is shown at the 1σ contour level. Residues 1-10 are drawn with thick lines and the electron density in represented by thin lines. The Glu1 and Asp2 residues adopt an α-helical conformation. Electron density corresponding to the $^{f}Met_{-1}$ residue is not observed; this residue has not been included in the rONC structure. The density associated with the Glu1 sidechain is the weakest in the structure. Lys9 and His10 make hydrogen bonded contacts to the active site sulfate anion through their respective side-chain atoms.

for the rONC structure. The bound sulfate anion does not have distinct lobes of electron density corresponding to the four oxygen atoms in maps calculated with either $3|F_o|-2|F_c|$ or $2|F_o|-|F_c|$ coefficients at this resolution. In omit maps (Bhat, 1988) calculated after refinement without the anion, the sulfate anion oxygen positions are observed in the electron density and correspond to the refined sulfate oxygen coordinates. The $|F_o|-|F_c|$ α_{calc} omit map electron density contoured at 3σ is shown in Figure 4.2.

The Ramachandran plots and average B-factor ($B = 8\pi\overline{u^2}$, where $\overline{u^2}$ is the mean-square vibrational amplitude) versus sequence number plots for rONC and ONC (Mosimann *et al.*, 1994) are virtually identical and have not been provided. Only Asp2 has appreciably different ϕ, ψ dihedral angles in the Ramachandran plots of the two structures. Asp2 has β-strand like ϕ (-160°) and ψ (169°) angles in ONC but adopts an α-helical conformation ($\phi=-77^{\circ}$, $y= 44^{\circ}$) in rONC. While the average main chain and side chain B-factor versus sequence number plots are similar for the two structures, the rONC average B-factors are roughly 25% larger than the equivalent average B-factor in the ONC structure. The one large divergence in the average B-factor (<B>) plots occurs at the N-terminus: ONC:Pyr1 main-chain <B>=19.9Å^2, side-chain <B>=22.7Å^2, rONC:Glu1 main-chain <B>=49.7Å^2, side-chain <B>=64.8Å^2. The absence of electron density corresponding to $^{f}Met_{-1}$ and the high B-factors associated with Glu1 indicate the N-terminus of rONC is more mobile than its counterpart in ONC.

Molecular conformation

The rONC structure can be superimposed on the ONC structure with an rmsd of 0.69Å for all 104 C^{α} atoms and 0.18Å for the 103 C^{α} atoms of residues Asp2-Cys104 (Figure 4.3). The secondary structural elements (helix α1 to α3, strand β1 to β6) described in the ONC structure are present here with one significant difference. In rONC, helix α1 is two residues longer and extends from Glu1 to His10. The conformational difference in Asp2 adds this residue and Glu1 to helix α1 in rONC. As a result,

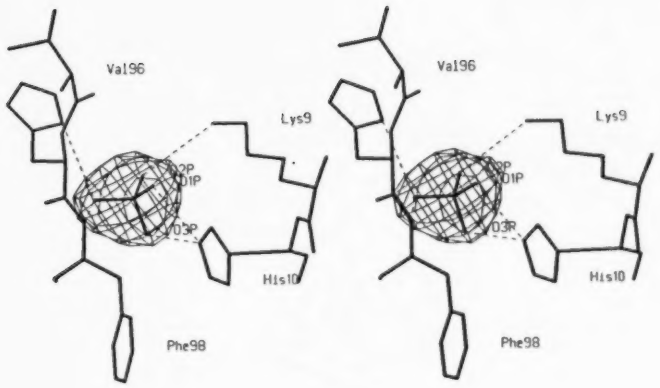


Figure 4.2 The $|F_o|-|F_c|\ \alpha_{calc}$ electron density associated with the sulfate anion site after refinement leaving out the atoms of the anion. The electron density is contoured at the 3σ level. The density is roughly spherical and does clearly indicate four distinct lobes of electron density corresponding to the sulfate oxygens.

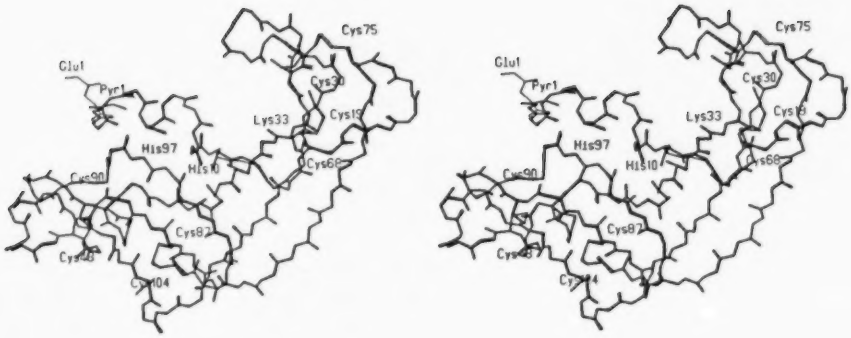


Figure 4.3 A representation of the main chain atoms of the least-squares superposed structures of ONC (thin lines) and rONC (thick lines). The rmsd for 103 C^{α} atoms is 0.18Å. Significant conformational differences in the two structures are confined to the N-terminus. In the ONC structure the Pyr1 and Asp2 residues fold back against the N-terminal α-helix. In rONC the Glu1 and Asp2 residues continue in an α-helical conformation. The C^{α} atoms of Pyr1 and Glu1 are separated by 7Å.

the C^{α} and C^{δ} atoms of Pyr1 (ONC) and Glu1 (rONC) are separated by 7.0 and 8.7 Å, respectively. This leads to the large rmsd for the 104 C^{α} atoms of these two structures.

Figure 4.4 is a stereo diagram of the superposed active sites of rONC and ONC and Table 4.2 lists the hydrogen bonds involving the sulfate anion oxygens. Other than the conformational differences of the two N-terminal residues, the active sites of the two structures are virtually identical. Lys9N$^{\zeta}$, His10N$^{\varepsilon 2}$, His97N$^{\delta 1}$ and Phe98N occupy equivalent positions in the least-squares superposition of these structures and the sulfate anions are bound in nearly identical orientations. The sulfur atom positions of the superposed anions are only slightly different (0.29Å). In both structures, His10N$^{\varepsilon 2}$ forms a bifurcated hydrogen bond with the sulfate oxygens O1 (SulO1) and O3 (SulO3) whereas His97N$^{\delta 1}$ and Phe98N donate hydrogen bonds to SulO4. In ONC, Lys9N$^{\zeta}$ donates hydrogen bonds to SulO2 (3.09Å) and Pyr1O$^{\varepsilon 1}$ (2.91Å). In rONC there is no structural equivalent to Pyr1 and the Lys9N$^{\zeta}$ to SulO2 hydrogen bond distance is increased by 0.4Å. Wat254 and Wat 258 of rONC hydrogen bond to SulO2 and SulO1, respectively and have counterparts in ONC. There is an additional intermolecular hydrogen bond from K55'N$^{\zeta}$ of a symmetry related molecule at -x+1/2, -y, z-1/2 to SulO4 in both structures. The K55' side chain is not expected to perturb the conformation of the active site very much (Mosimann *et al.*, 1994).

There are 72 and 96 ordered solvent molecules in the rONC and ONC structures, respectively. A total of 54 of the 72 ordered solvent molecules are common (within 1.0Å) in the two structures (rmsd=0.48Å). The difference in the number of ordered solvent molecules likely results from two sources, differences in the resolution (2.2Å versus 1.7Å for ONC) and in the refinement technique. The rONC structure has been refined with TNT release 5D (Tronrud, 1992) using a solvent continuum whereas ONC was refined using PROLSQ (Hendrickson & Konnert, 1981) and no solvent continuum.

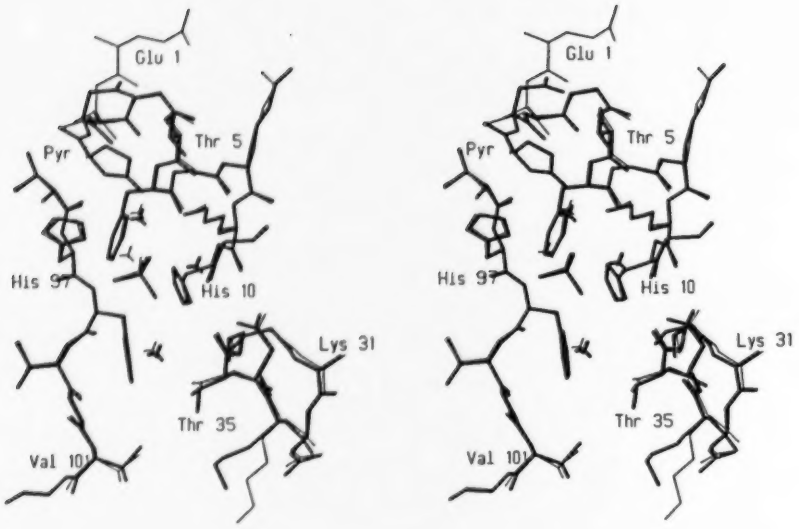


Figure 4.4 An all atom representation of the active sites of the least-squares superposed structures of ONC (thick lines) and rONC (thin lines). Ordered solvent molecules (small tetrahedra) and the bound sulfate anion observed in each structure are also included. The active site residues Lys9, His10, His97 and Phe98 have nearly identical conformations in the two structures. The sulfate anion is bound in the same orientation with a slightly shifted position. The major conformational differences in the two structures are confined to the N-terminus (Glu1 and Asp2).

Table 4.2

Conserved hydrogen bonds involving the sulfate anion in the superposed ONC and rONC structures

Protein Atom name+	Sulfate Atom name	distance(Å) ONC	distance(Å) rONC
Lys9Nζ	O2	3.09	3.51
His10Nε2	O1	2.69	2.88
His10Nε2	O3	3.32	3.05
His97Nδ1	O4	2.93	3.08
Phe98N	O4	3.20	3.03
Wat254	O2	3.02	2.99
Wat258	O1	2.51	2.50
Lys55'Nζ‡	O4	2.79	2.82

+ Wat numbers are taken from rONC. In ONC, Wat405 and Wat306 are equivalent to Wat 254 and Wat258.

‡ Lys55' is from a symmetry related molecule and forms an intermolecular contact.

Superpositions of rONC and ONC with RNase A

The rONC and ONC structures have been superposed by least-squares methods with RNase A coordinates (1RPF, 1RPG, 1RPH; Zegers *et al.*, 1994) obtained from the Protein Data Bank (Bernstein *et al.*, 1977). The rmsds for the overlaps of rONC and ONC with RNase A (1RPG) are 3.16Å and 3.22Å, respectively for the 102 equivalent C^{α} atoms. Alternatively, the rmsds for the C^{α} atoms that can be superposed within 3.8Å are 1.37Å for rONC (89 C^{α}) and 1.31Å for ONC (87 C^{α}). The N-terminal residue and 14 residues from loop structures in ONC (15-18, 24-27, 49-54) account for the 15 residues whose C^{α} atoms do not superpose within 3.8Å of equivalent RNase A atoms.

The ONC and rONC residues that comprise the anion binding site come from helix-α1 (rONC:Glu1-His10, ONC:Trp3-His10) and strand-β6 (Ala94-Val101). In the ONC superposition with RNase A based upon the 87 C^{α} atoms that superpose within 3.8Å, the secondary structures contributing residues to the active site are shifted relative to their counterparts in RNase A. In this superposition the N, C^{α}, C, O and C^{β} atoms of the ONC active site residues (Lys9, His10, His97 and Phe98) and their counterparts in RNase A superpose with an rmsd of 0.62Å. The difference in position of these residues is a result of a shift observed in the position of the two lobes of ONC (lobe 1: helices-α1, α3, strands-β2, β5, β6; lobe 2: helix-α2, strands-β1, β3, β4) when compared to RNase A. The 52 C^{α} atoms of lobe 1 of ONC (residues 1-12, 41-62, 85-102) have been least-squares superposed with the equivalent C^{α} atoms of RNase A to improve the overlap of the active site residues. The rmsd is 0.80Å for the 45 out of 52 C^{α} atoms of lobe 1 that can be superposed within 3.8Å and the main chain and C^{β} atoms the ONC and RNase A active site residues have an rmsd of 0.32Å in this improved superposition. Figure 4.5 shows the least-squares superposition of ONC and RNase A using the 45 of 52 C^{α} atoms of lobe 1 that can be superposed within 3.8Å. In Figure 4.5, the ONC and RNase A residues of lobe 2 are rotated slightly relative to one another as a result of the improved

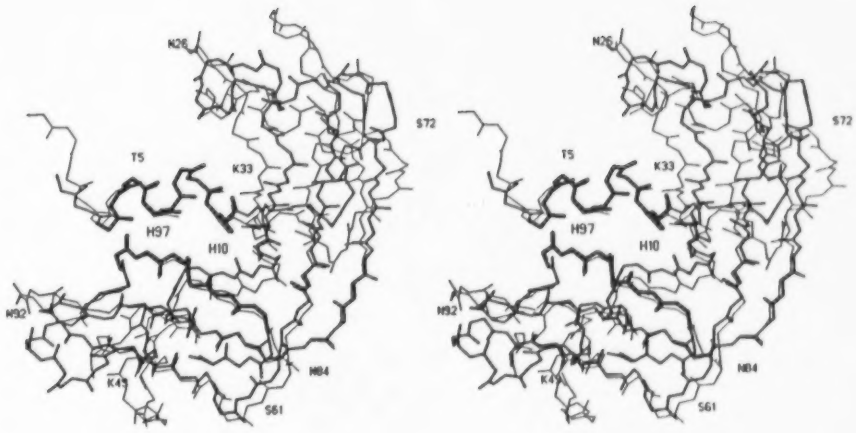


Figure 4.5 The least-squares superposition using 52 C^{α} atoms of lobe 1 of rONC(1-12, 41-62 and 85-102; thick lines) and the equivalent atoms in RNase A (1RPG; thin lines). The rmsd is 0.80Å for 45 out of 52 C^{α} coordinates that can be superposed within 3.8Å. The helix-α1 and strand-β6 of lobe 1 contain the active site residues in these structures and are shown in heavy lines. The rmsd between the superposed main-chain and C^{β} atoms of the active site residues of ONC (Lys9, His10, His97 and Phe98) and RNase A is 0.32Å in the superposition based on lobe 1. The rmsd is 0.62Å for same active site atoms when the superposition based on the entire molecule.

agreement in the overlap of the active site residues of lobe 1. The angles between strands-β1, β3, β4 of lobe 2 of ONC and their counterparts in RNase A are 7.9°, 7.4° and 7.5° respectively.

The position and orientation of the sulfate anion bound to ONC has been compared to the sulfate anion bound to RNase A. The superposed ONC and sulfate containing RNase A (1RPH; Zegers *et al.*, 1994) active sites are shown in Figure 4.6 using the improved superposition. The sulfate anions bind in closely similar positions and orientations in the two structures and suggest that ONC and RNase A will bind substrate in an equivalent manner. The conserved His10 and Phe98 residues contact equivalent anion oxygens in the two structures as does Lys9N$^{\zeta}$ (ONC) and its counterpart Gln11 (RNase A). His97 has a side chain conformation (χ_1=-73°, χ_2=-53°) that is different from that of the equivalent RNase A residue (χ_1=153°, χ_2=80°) in the complex. The ONC His97 side chain conformation is similar to one of two His119 conformations observed in several uncomplexed RNase A structures (Borkakoti *et al.*, 1982; Martin *et al.*, 1987; deMel *et al.*, 1992). In both the ONC and the rONC structures, the Lys55N$^{\zeta}$ intermolecular contact with the sulfate anion sterically prevents His97 from adopting the side chain conformation observed in the RNase A structure. In solution, His97 will be able to adopt a conformation equivalent to that seen in the RNase A. The invariant Lys31 (ONC) and Lys41 (RNase A) C$^{\alpha}$ atoms are 3.04Å apart in the superposition.

In order to deduce a possible substrate binding mode, ONC-d(UpG) and rONC-d(UpG) model structures have been constructed based upon the RNase A-d(CpA) complex structure. To generate the models, the individual lobes of ONC were superposed with the RNase A-d(CpA) complex and the d(CpA) inhibitor was replaced by d(UpG), an analog of the preferred substrate for ONC. In the RNase A reaction mechanism, the ribose 0(2') is the nucleophile in the initial transesterification that produces a cyclic-2',3'-monophosphate intermediate and 0(2') is the leaving group in the subsequent hydrolysis to

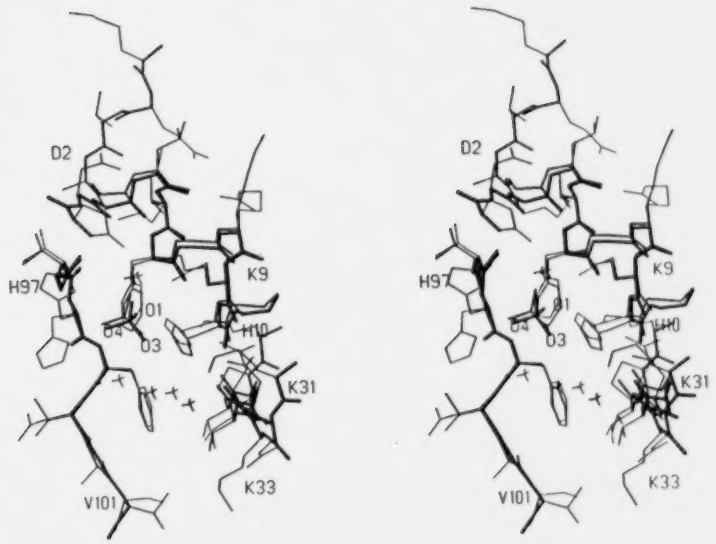


Figure 4.6 The superposed ONC and sulfate containing RNase A (1RPH) active sites using the superposition from Figure 4.5. The ONC structure is shown as bold lines and RNase A as light lines. Main chain atoms are represented by thicker lines than side chain atoms. The position and orientation of the bound sulfate anion of ONC and RNase A are similar. The active site histidines residues contact equivalent anion oxygens in the two structures as do Lys9 (ONC) and its counterpart Gln11 (RNase A). The His97 conformation in ONC corresponds to an alternate conformation observed in other RNase A structures.

a 3'-monophosphate product. d(CpA) is a non-hydrolysable deoxy form of a diribonucleotide substrate that lacks the ribose 0(2'). The ONC-d(UpG) and rONC-d(UpG) model structures are seen in Figures 4.7a-b as well as suggested hydrogen bonds to the inhibitor. The sulfur atoms of the ONC and rONC anions are 0.57Å and 0.62Å from the superposed position of the phosphate atom of the d(CpA), respectively. An equivalent superposition of sulfate containing RNase A and the RNase A-d(CpA) complex, the comparable distance is 0.31Å.

The proposed hydrogen bonds in the ONC-d(UpG) modeled structure have been compared to the hydrogen bonds in the RNase A-d(CpA) structure. The base, ribose and phosphate subsites are referred to as B_1, R_1, P_1, B_2 and R_2, respectively (Richards & Wykoff, 1971). In the ONC-inhibitor model, the hypothetical hydrogen bonds from Lys9N$^{\zeta}$ and His10N$^{\varepsilon 2}$ to the phosphate moiety of the inhibitor (P_1) are of reasonable length (2.81Å and 3.20Å). The His10N$^{\varepsilon 2}$ hydrogen bond is similar to that seen in the RNase A complex structure and the Lys9N$^{\zeta}$ hydrogen bond is analogous to RNase A's Gln11N$^{\varepsilon 2}$ hydrogen bond. Phe98N makes a 2.80Å hydrogen bond with the phosphate oxygen equivalent to SulO1. By adopting the conformation of the equivalent RNase A residue (χ_1=153°, χ_2=80°), His97 (ONC) can be brought into hydrogen bond distance of the phosphate (2.62Å). Likewise, the invariant Lys31 (ONC) side chain is modeled with an all-*trans* conformation as observed for Lys41 (RNase A) and is within 3.2Å of the O(3') ribose of uridine.

In Figure 4.7, the uracil base O(2) and N(3) atoms are paired with the main chain N and side chain O$^{\gamma 1}$ atoms of the invariant Thr35 (Thr45 in RNase A) in the B_1 subsite. These hydrogen bonds are 3.06Å and 2.43Å long in the ONC-d(UpG) model. The O(4) atom of the uracil base is directed towards Ser123 in RNase A. The equivalent residue in ONC is Val101. The nearest potential protein hydrogen bond acceptor or donor is Lys33N$^{\zeta}$ and it is 3.06Å from O(4) when Lys33 adopts an all-*trans* side chain

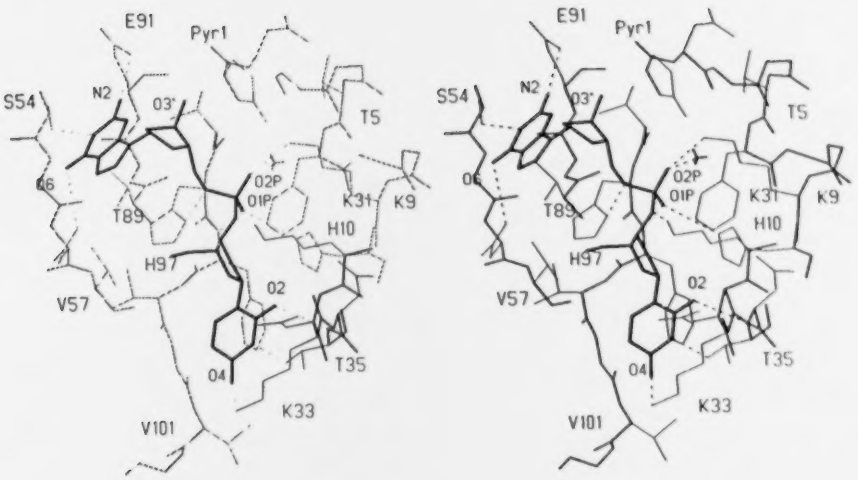


Figure 4.7a The active site of the ONC-d(UpG) model based on RNase A-d(CpA) . The ONC side chains are shown as thin lines, the main chain as thicker lines and d(UpG) is shown as thick lines. Potential hydrogen bonds between ONC and d(UpG) are indicated as dashed lines. The ONC active site residues Lys9, His10, His97 and Phe98 make contacts equivalent to those observed in the RNase A-complex structure. Adjustment of the Lys33 side chain torsion angles brings Lys33N$^{\zeta}$ within hydrogen bonding distance of the uridine O4. The guanine base has three nearby hydrogen bonding partners; Ser54O$^{\gamma 1}$ (N1), Asn56N$^{\varepsilon 2}$ (O6) and Glu91O$^{\varepsilon 1}$ (N2).

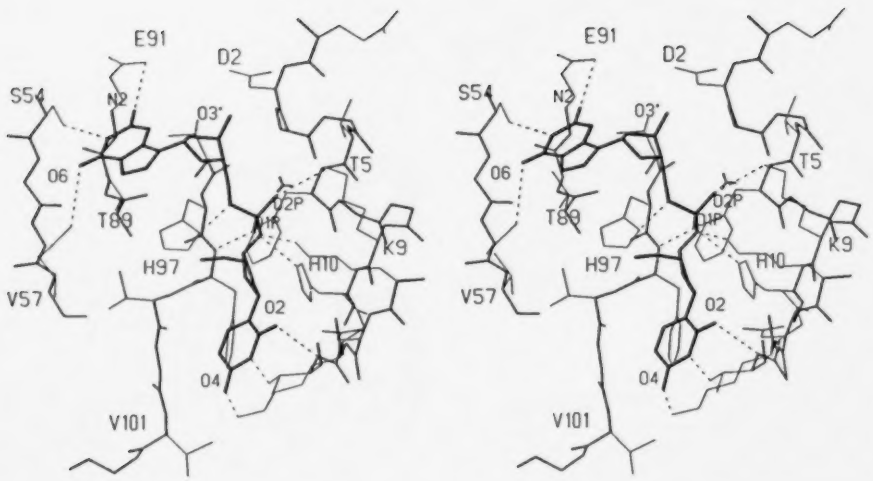


Figure 4.7b The rONC-d(UpG) model is shown in a manner equivalent to 7a. The guanosine deoxyribose packing interaction with Pyr1 (Figure 7a) is partially replaced by the Asp2 carboxylate in the rONC model. This change in the shape and charge of the ribose binding pocket may be responsible for rONC's lack of significant ribonucleolytic activity.

conformation. The guanine binding site (B_2) in ONC is considerably different from that in RNase A as a result of a five residue insertion in RNase A (65-69) relative to ONC. In the RNase A-d(CpA) complex, adenine N(1)and N(6) atoms are paired with Gln69 and Asn71 in the B_2 binding site. In the ONC-d(UpG) model the N(1) and O(6) atoms of guanine are within hydrogen bond distance of Ser54$O^{\gamma 1}$ (2.7Å) and Asn56$N^{\delta 2}$ (3.5Å). Structure based sequence alignments indicate Ser54 and Asn56 are equivalent to Thr70 and Cys72 (RNase A) (Mosimann *et al.*, 1994). ONC Glu91$O^{\varepsilon 1}$ is 2.6Å from the N(2) position of the base and provides a third hydrogen bond in the model. The conserved Glu111$O^{\varepsilon 1}$ is the equivalent residue in RNase A.

DISCUSSION

Significant main chain differences in the rONC and ONC protein structures are confined to the N-terminus. In rONC, the N-terminal helix (Glu1-Trp10) extends into a solvent channel. The ^{f}Met$_{-1}$ residue is disordered and not seen in the electron density maps. The high B-factors and relatively weak density associated with Glu1 sidechain suggests a high degree of mobility and are consistent with ^{f}Met$_{-1}$ being disordered. In ONC, the N-terminal helix ends at Trp3 and Pyr1 is folded back against this helix and the cyclized pyroglutamyl residue is directed into the active site.

The positional and conformational differences in the anion binding residues in rONC and ONC are small and the bound sulfate anion is only slightly shifted (0.29Å between sulfur atoms) in the two structures. In ONC, Pyr1O^{ε} forms hydrogen bonds with Lys9N^{ζ} (2.91Å) and Wat306 (2.66Å), respectively. While rONC does not have a structural counterpart to Pyr1, the Lys9N^{ζ} and Wat258 positions differ by less than 0.5Å in the superposition. This suggests the anion binding residues of rONC are correctly positioned for catalysis. In the absence of additional structural information, there is no obvious structural reason for the rONC enzymatic inactivity. The ONC-d(UpG) and rONC-d(UpG)

models may explain rONC's inactivity. In the ONC-d(UpG) model, the guanine ribose O(3') is 3.9Å from Pyr1Cβ as the ribose packs against the Pyr1 ring in the R_2 subsite. The C(2')-exo conformation of the ribose and the *anti* conformation of the glycosyl torsion angle allow the purine base to interact with up to three hydrogen bonding partners in the B_2 subsite. In the rONC-d(UpG) model the wan der Waals contacts to the ribose are lost and are only partially replaced by the Asp2 carboxylate. Maximizing the non-bonded contacts in the R_2 subsite of rONC brings O(3') nearer to the Asp2 negative charge. Since the O(3') of the ribose of guanosine forms an ester bond with the anionic phosphate group in oligoribonucleotide substrates, an unfavorable electrostatic interaction may result. In addition, Asp2O$^{\delta 2}$ is already 5.31Å from the Glu91 γ-carboxylate that hydrogen bonds to N(2) of guanine in the model. Large conformational changes involving the ribose to improve van der Waals interactions or relieve electrostatic interactions may move the purine base from its binding pocket (B_2). Together, the change in the shape and electrostatic nature of the ribose binding pocket of guanosine (R_2) may give rise to rONC's lack of activity towards extended substrates. It is also possible the mobility of the N-terminus of rONC contributes to its lack of activity.

ONC has a k_{cat}/k_m of 4.73 x 10^3 $M^{-1}sec^{-1}$ for UpG diribonucleotide substrates at its optimal pH of 6 (Boix *et al.*, unpublished results). Under the same conditions, RNase A has a k_{cat}/k_m of 2.93 x 10^5 $M^{-1}sec^{-1}$ for the same substrate. ONC has a marked preference for UpG but it will slowly hydrolyze CpG as well (W. Ardelt, unpublished results). In contrast, RNase A cleaves UpX and CpX (X = A,G,C,T) diribonucleotides and is least-active against UpG (Harper & Vallee, 1989). The superposition of ONC with the sulfate containing RNase A structure shows that ONC and RNase A bind the active site anion in equivalent positions. In the RNase A-d(CpA) structure and in the ONC-d(UpG) model, the N and O$^{\gamma 1}$ atoms of the invariant Thr (ONC Thr35; RNase A Thr45) form hydrogen bonds with the O(2) and N(3) atoms of the pyrimidine base (B_1). This suggests that differences in the pyrimidine-base specificity of these two enzymes arise from differences in the binding of

the N(4) (cytosine)/O(4) (uracil) atom. In the RNase A-inhibitor structures the N(4) (cytosine) or O(4) (uracil) atom is directed at Ser123O$^{\gamma 1}$. Direct hydrogen bonds between N(4) and Ser123O$^{\gamma 1}$ (Richards & Wykoff, 1971) and water-mediated interactions between O(4) and Ser123O$^{\gamma 1}$ (Wlodawer *et al.*, 1983; Borah *et al.*, 1985) have been reported. Since Ser123O$^{\gamma 1}$ and water molecules can either accept or donate hydrogen bonds, they account for the pyrimidine base preference of RNase A.

In the ONC-d(UpG) model (Figure 4.7) the uracil O(4) hydrogen-bonding potential is unsatisfied. There are three ONC atoms within 6Å of the uracil O(4) that are capable of participating in hydrogen bonds: Lys33N$^{\zeta}$, Thr35O$^{\gamma 1}$ and Val101N. From previous crystallographic work, the invariant Thr35O$^{\gamma 1}$ atom is known to interact with the N(3) position of pyrimidine bases (B_1). Val101N is 4.11Å away and the conformation of this residue is fixed by the neighboring Cys87 to Cys104 disulfide bridge. In ONC, Lys33 and Val101 are equivalent to Val43 and Ser123 of RNase A. Adjustment of the Lys33 side chain torsion angles can bring the N$^{\zeta}$ atom to within 3.1Å of the uridine O(4) atom while retaining good stereochemistry. After adjusting the Lys33 torsion angles, the angle between C(4), O(4) (uridine) and Lys33N$^{\zeta}$ is 127° which is close to the expected 120° hydrogen bonding angle. An interaction between Lys33N$^{\zeta}$ and N4 (cytosine)/O4 (uracil) provides an attractive explanation of the observed B_1 site specificity of ONC. As Lys33N$^{\zeta}$ is only capable of donating hydrogen bonds at physiologically relevant pH, it would strongly favor the uridine O(4) atom (hydrogen bond acceptor) over the cytosine N(4) atom (hydrogen bond donor). The guanine specificity of ONC in the B_2 site can also be explained by the nature of the residues in this subsite. Ser54 is 2.5Å from the guanine N(2) and can form hydrogen bonds with the N(2) atom of either purine base. In both ONC and rONC, Asn56N$^{\delta 2}$ hydrogen bonds with Thr89O$^{\gamma 1}$ (2.66Å, 2.56Å). In this conformation, Asn56N$^{\delta 2}$ is available to interact with O(6) of guanine but not the N(6) of adenine. Similarly, the Glu91 carboxylate can productively interact with guanine N(2) (2.5Å) but not with adenine. The Asn56N$^{\delta 2}$ and Glu91O$^{\varepsilon 1}$ interactions are proposed to

account for the guanine specificity of ONC.

The difference in the relative positions of the RNase A and ONC lobes (see Results) suggests the need for a conformational change in ONC during substrate binding and catalysis. The superposition of ONC and RNase A based on 45 out of 52 equivalent C^{α} atoms of lobe 1 that can be overlapped within 3.8Å is justified by the improved agreement in the superposed positions of the active site residues of ONC and RNase A. The rmsd for the main chain and C^{β} atoms of the active site residues (Lys9, His10, His97, Phe98) drops from 0.62Å to 0.32Å in superpositions based on the entire structure and the atoms of lobe 1, respectively. In Figure 4.5, lobe 2 of ONC is rotated relative to lobe 2 of RNase A and the active site cleft of ONC is more exposed. As a consequence, the Lys31 of ONC is 2.0Å (C^{α}-C^{α} distance) from the required Lys41 of RNase A. In the RNase-d(CpA) complex, the Lys41 side chain adopts an all-*trans* conformation and hydrogen bonds to O(3') in the R_1 subsite. Lys31 of ONC is over 4Å from O(3') and must undergo a relatively large conformational change in order to make an equivalent contact. The difference in the lobe 2 positions of ONC and RNase A in Figure 4.5 is a hinge-like closing of approximately 8° (see Methods). In Figure 4.7, lobe 1 and lobe 2 of ONC were separately superposed with RNase A. In the resulting ONC-d(UpG) model, Lys31 has a position equivalent to Lys41 of RNase A and Lys33 is brought closer to the B_1 subsite. The closer approach of Lys33N^{ζ} to the uridine O(4) position supports the interpretation that Lys33N^{ζ} contributes to the pyrimidine specificity of ONC. Finally, this type of conformational change has not been observed in complexed and uncomplexed RNase A structures. In the recent crystallographic structure of angiogenin, the pyrimidine binding site (B_1) is sterically blocked by the Glu117 side chain (Acharya, *et al.*, 1994). As a consequence, substrate binding requires a conformational change that moves Glu117 and adjacent residues. Within the RNase superfamily, angiogenin also has a relatively small RNase activity (2%) in comparison with RNase A (Shapiro *et al.*, 1986).

The authors realize that the ONC-d(UpG) and rONC-d(UpG) models developed

above cannot replace direct structural studies of inhibitor complexes. These models suggest differences in the shape and electrostatic nature of the R_2 result in the lack of rONC catalytic activity. In addition, the modeling has identified residues of the novel B_2 subsite and provided chemically reasonable explanations for the observed substrate specificity.

MATERIALS AND METHODS

Protein Purification and Crystallization

A synthetic rONC gene having an N-terminal sequence $^{f}Met_{-1}Glu_1Asp_2\cdots$ has been expressed in an *Escherichia coli* expression system and the resulting protein has been purified from cell lysates (Wu, Y.N. & Youle, R.J., unpublished results). The identity of the amino-terminal residues has been confirmed by determining the amino acid composition and by N-terminal amino acid sequencing (W. Ardelt, unpublished data).

The purified rONC was dissolved in water (10 mg/mL; pH 6.0) and screened against a set of crystallization conditions, based upon a sparse matrix sampling method (Jancarik & Kim, 1991), using hanging drops. rONC was also screened against the crystallization conditions that produced native ONC crystals (Mosimann *et al.*, 1992) and small rosettes were obtained. Ultimately, the largest single crystals (up to 0.15 mm x 0.12 mm x 0.5 mm) were grown over a 20% w/v polyethylene glycol (PEG) 1000, 0.2 M $(NH_4)_2SO_4$, 0.010 M acetate solution at pH 4.5. The initial hanging drop conditions contained 4 μL of 10mg/mL rONC and 6 μL of reservoir solution at pH 4.5.

Data collection

The unit cell and space group of the rONC crystals were determined from X-ray diffraction patterns and found to be isomorphous with the native ONC crystals. Data were collected by exposing the crystal to $CuK_{\bar{\alpha}}$ radiation (λ=1.5418Å) produced by a Rigaku RU-200 rotating anode X-ray source operated at 40 kV and 150 mA. The X-ray diffraction

intensity data were measured on a San Diego Multiwire Systems (SDMS) multiwire proportional counter (Hamlin *et al.*, 1985) while the crystal was moved through a series of ω-motor scans and ϕ, χ settings determined using the program LATTICEPATCH (Klinger & Kretsinger, 1989). The measured intensity data were corrected with Lorentz and polarization factors, scaled and multiple measurements of individual reflections were merged and averaged using the SDMS software. The data were subsequently placed on an absolute scale. Table 4.3 lists the unit cell parameters, the completeness of the collected data and the agreement among symmetry equivalent structure factor amplitudes for the 2.2Å resolution data set.

Refinement

The initial phases for the rONC structure were derived from the native ONC coordinates (1ONC; Mosimann *et al.*, 1994) using residues 2-104. The initial rONC model was refined against the observed structure-factor amplitudes ($|F_o| \geq 1\sigma|F_o|$) between 10Å and 2.2Å. resolution using the simulated annealing slow-cooling protocol described in the X-PLOR manual (A. Brünger *et al.*, 1987). This individual B-factor ($B= 8\pi^2\overline{u^2}$, where $\overline{u^2}$ is the mean-square vibrational amplitude) refinement and positional parameter refinement resulted in a conventional crystallographic R-factor ($R=\Sigma||F_o|-|F_c||/\Sigma|F_o|$, where F_o and F_c are the observed and calculated structure factor amplitudes) of 0.213. Subsequent refinement was performed with TNT version 5C and 5D (Tronrud, 1992) using data between 20.0Å and 2.2Å resolution with $|F_o| \geq 1\sigma|F_o|$. Electron density maps were calculated using SIGMAA $2mF_o$-DF_c and mF_o-DF_c coefficients (Read, 1986) and using TNT $3|F_o|$-$2|F_c|$, $2|F_o|$-$|F_c|$ and $|F_o|$- $|F_c|$ coefficients. Interactive graphics refitting sessions were performed using TOM 3.0 Alberta/CalTech (based upon Jones, 1985). Throughout the refinement procedure the stereochemical quality of the rONC model was monitored using PROCHECK (Laskowski, 1993).

Table 4.3

Summary of data collection statistics for Onconase(Met-1,Glu1) native data set[+]

	Resolution (Å)	No. Ref.	Unique Ref.	% Complete	I/σ(I)	R_m[‡]	unit cell (Å) *a*	*b*	*c*
Native	2.2	36332	5066	98.0	12.0	0.053	41.14	69.37	32.53

[+] Space Group $P2_12_12_1$; wavelength ($CuK_{\bar{\alpha}}$) = 1.5418Å

‡ $R_m = \Sigma_{hkl}\Sigma|I_{obs} - I_{ave}|/\Sigma|I_{ave}|$, where I_{obs} and I_{ave} are the observed intensity of a single reflection and the average intensity of all measurements of a reflection and its Freidel mate, respectively.

A sulfate ion bound in the active site region and the Glu1 residue were fit to the electron density maps calculated after the initial TNT refinement and Asp2 was refit. There was no electron density corresponding to $^{f}Met_{-1}$ and this residue has not been included in the model. Ordered solvent molecules were added during interactive graphics refitting sessions. Solvent molecules were required to meet three conditions before being selected: (1) the $|F_o|-|F_c|$ electron density map had a peak greater than 2.5σ at the proposed solvent site, (2) the electron density was roughly spherical in shape and (3) a productive hydrogen bond was formed with a protein atom (or with an already established ordered solvent molecule). All solvent molecules were subsequently refined at unit occupancy and were required to have refined B-factors less than $80Å^2$ (approximately 1.2 times the value of the largest B-factor of a protein atom). The electron density of the Glu1 side-chain remained discontinuous throughout the course of the refinement and electron density corresponding to $^{f}Met_{-1}$ was never observed. Prior to the final refinement cycles a model omitting Glu1 was subjected to a short refinement and the resulting model was used to calculate electron density maps (Bhat, 1988). The electron density in the region of Glu1 was not significantly different from that observed in electron density maps calculated with the complete model.

The final refined model of rONC contains 828 protein atoms, a sulfate anion and 72 ordered solvent molecules. As in the refinement of ONC, the anion was presumed to be SO_4^{2-} based upon its inclusion in the crystallization media (0.2 M $(NH_4)_2SO_4$) and the absence of added phosphate to those solutions. The refined coordinates of several water molecules with high B-factors were fixed during the latter stages of refinement. These solvent molecules have refined hydrogen bond distances of slightly less than 2.5Å. Given the relatively weak electron density associated with these solvent molecules, they have been fixed at a hydrogen bonding distances of 2.5Å or greater.

Three amino acid side-chains have been modeled with alternate conformations. Ser 54 is modeled in two conformations while Thr60 and Thr83 are modeled in three

conformations. The occupancies of the various alternate conformations were fixed at 0.33 or 0.5 and each alternate conformation had to improve the fit between the model and the resulting electron density. These residues are all located on the surface of the rONC structure and extend into solvent channels.

All least-squares superpositions were performed using O (Jones *et al.*, 1991). The superposition of rONC and ONC was performed using the C^{α} coordinates for residues 2-104. Superpositions involving RNase A (1RPG) made use of the published amino acid sequence alignment (Mosimann *et al.*, 1992) with the modifications described in Mosimann *et al.*, 1994. The superposition in Figure 4.5 was performed based on the 52 equivalent C^{α} atoms of lobe 1 of ONC (residues 1-12, 41-62 and 85-102) and RNase A. In this superposition, lobe 2 of ONC is rotated relative to lobe 2 of RNase A. The angle between the antiparallel β-sheet (strands β1, β3 and β4) of lobe 2 of ONC and RNase A is approximately 8°. For strands-β1 and β4, the angle has been measured from using the N-terminal residue of the rONC strand, the C-terminal residue of the ONC strand and the N-terminal residue of the RNase A strand. For strand-β2, the angle has been measured from using the C-terminal residue of the rONC strand, the N-terminal residue of the ONC strand and the C-terminal residue of the RNase A strand. In ONC-d(UpG) model in Figure 4.7, lobe 1 and lobe 2 of ONC (residues 13-40 and 63-84) have been separately superposed with their counterparts from RNase A.

REFERENCES

Acharya KR, Shapiro R, Allen SC, Riordan JF, Vallee BL. 1994. Crystal structure of human angiogenin reveals the structural basis for its functional divergence from ribonuclease. *Proc. Natl. Acad. Sci. U.S.A. 91*:2915-2919.

Bernstein FC, Koetzle TF, Williams GJB, Meyer Jr. EF, Brice MD, Rodgers JR, Kennard O, Shimanouchi T, Tasmui M. 1977. The Protein Data Bank: A computer-based archival file for macromolecular structures. *J. Mol. Biol. 112*:535-542.

Bhat, TN. 1988. Calculation of an OMIT map. *J. Appl. Crystallog.* **21**, 279-281.

Beintema JJ, Schuller C, Irie M, Carsana A. 1988. Molecular Evolution of the ribonuclease superfamily. *Progr. Biophys. Mol. Biol. 51*:165-192.

Borah B, Chen C, Egan W, Miller M, Wlodawer A, Cohen JS. 1985. Nuclear magnetic resonance and neutron diffraction studies of the complex of ribonuclease A with uridine vanadate, a transition state analogue. *Biochemistry 24*:2058-2067.

Borkakoti N, Moss DS, Palmer RA. 1984. The refined structure of ribonuclease-A at 1.45Å resolution. *Acta Cryst B38*:2210-2217.

Carlisle CH, Palmer RA, Mazumdar SK, Gorinsky BA, Yeates DGR. 1974. The structure of ribonuclease at 2.5Å resolution. *J. Mol. Biol. 85*:1-18.

Darzynkiewicz Z, Carter SP, Mikulski SM, Ardelt W, Shogen K. 1988. Cytostatic and cytotoxic effects of Pannon (P-30 protein), a novel anticancer agent. *Cell Tissue Kinet. 21*:169-182.

deMel VSJ, Martin PD, Doscher MS, Edwards BFP. 1992. Structural changes that accompany the reduced catalytic efficiency of two semisynthetic ribonuclease analogues. *J. Biol. Chem. 267*: 247-256.

Harper JW, Vallee BL. 1989. A covalent angiogenin/ribonuclease hybrid with a fourth disulfide bond generated by regional mutagenesis. *Biochemistry 24*:2058-2067.

Hendrickson WA, Konnert JH. 1981. In: Srinivasan, R. *Biomolecular Structure, Function, Conformation and Evolution.* Vol. 1. Oxford: Pergamon. pp. 43-57.

Jancarik J, Kim SH. 1991. Sparse matrix sampling: a screening method for crystallization of proteins. *J. Appl. Cryst. 24*:409-411.

Jones TA. 1985. Interactive computer graphics: FRODO. *Methods Enzymol. 115*:157-171.

Jones TA, Zou JY, Cowan SW, Kjeldgaard M. 1991. Improved Methods for building protein models in electron density maps and the location of errors in these models. *Acta Cryst. A47*:110-119.

Kabsch W, Sander C. 1983. Dictionary of protein secondary structure: pattern recognition of hydrogen bonded and geometrical features. *Biopolymers 22*:2577-2637.

Klinger AL, Kretsinger RH. 1989. LATTICEPATCH- an interactive graphics program to design data measurement strategies for area detectors. *J. Appl. Cryst. 22*:287-293.

Laskowski R. 1993. PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Cryst. 26*:283-291.

Martin PD, Doscher MS, Edwards BFP. 1987. The refined structure of a fully active semisynthetic ribonuclease at 1.8Å resolution. *J. Biol. Chem. 262*:15930-15938.

Mikulski SM, Ardelt W, Shogen K, Bernstein EH, Menduke H. 1990. Striking increase of survival of mice bearing M109 Madison carcinoma treated with a novel protein from amphibian embryos. *J. Natl. Cancer Inst. 182*:151-153.

Mikulski SM, Grossman AM, Carter PW, Shogen K, Costanzi JJ. 1993. Phase I human clinical trial of ONC^{R1} (P-30 protein) administered intravenously on a weekly schedule in cancer patients with solid tumors. *Int. J. Oncol. 3*:57-64.

Mosimann SC, Ardelt W, James MNG. 1994. Refined 1.7 Å X-ray crystallographic structure of P-30 protein, an amphibian ribonuclease with anti-tumor activity. *J. Mol. Biol. 236*:1141-1153.

Mosimann SC, Johns KL, Ardelt W, Mikulski SM, Shogen K, James MNG. 1992. Comparative molecular modeling and crystallization of P-30 protein: a novel anti-tumor protein of *Rana pipiens* oocytes and early embryos. *Prot. Structure Function and Genetics 14*:392-400.

Read RJ. 1986. Improved fourier coefficients for maps using phases from partial structures with errors. *Acta Cryst. A42*:140-149.

Richards FM, Wyckoff HW. 1971. The Enzymes (Boyer P. ed.) vol. 4:647-806, Academic Press, New York.

Shapiro R, Riordan JF, Vallee BL. 1986. Characteristic ribonucleolytic activity of human angiogenin. *Biochemistry 25*:3527-3532.

Titani K, Takio K, Kuwada M, Nitta K, Sakakibara F, Kawauchi H, Takayanagi G, Hakomori S. 1987. Amino acid sequence of sialic acid binding lectin from frog (*Rana catesbeiana)* eggs. *Biochemistry 26*:2189-2194.

Tronrud DE. 1992 Conjugate-direction minimization: an improved method for the refinement of macromolecules. *Acta Crsyt. A48*:912-916.

Wlodawer A, Miller M, Sjolin L. 1983. Active site of RNase: neutron diffraction study of a complex with uridine vanadate, a transition state analogue. *Proc. Natl. Acad. Sci. USA 80:*3628-3631.

Wu Y, Mikulski SM, Ardelt W, Ryback SM, Youle RJ. 1993. A cytotoxic ribonuclease: Study of the mechanism of onconase cytotoxicity. *J. Biol. Chem. 268:*10686-10693.

Zhao W, Beintema JJ, Hofsteenge J. 1994. The amino acid sequence of iguana (*Iguana iguana)* pancreatic ribonuclease. *Eur. J. Biochem. 219*:1875-1884.

Zegers I, Maes D, Dao-Thi M, Poortmans F, Palmer R, Wyns L. 1995. The structure of RNase A complexed with 3'CMP and d(CpA): Active site conformation and conserved water molecules. *Prot. Sci. 3*:2322-2339.

CHAPTER 5

X-RAY CRYSTALLOGRAPHIC STRUCTURE OF RECOMBINANT EOSINOPHIL-DERIVED NEUROTOXIN AT 1.83Å RESOLUTION†

INTRODUCTION

Eosinophil-derived neurotoxin (EDN) is a small, basic ribonuclease (RNase) of the eosinophil granule that is a member of the pyrimidine-specific RNase superfamily of vertebrates (Gleich *et al.*, 1986). It has been variably referred to as RNase U_S, RNase II and non-secretory RNase (Sierakowska & Shugar *et al.*, 1977) in the existing literature. Recombinant EDN (rEDN) is a neurotoxin that selectively kills cerebellar Purkinje cells (Gordon phenomenon; Durack *et al.*, 1981), is a potent cytotoxin when injected into *Xenopus* oocytes (Saxena *et al.*, 1991) and plays a prominent role in several human pathologies (Youle *et al.*, 1993 and references therein). When recombinant EDN (rEDN) is conjugated to monoclonal antibodies raised against a cell surface receptor, it is a potent and specific cytotoxin with favorable immunogenic properties (Ryback *et al.*, 1991). The homologous Onconase (Mikulski *et al.*, 1989), eosinophil cationic protein (ECP; Fredens *et al.*, 1982) and seminal RNase (D'Alessio *et al.*, 1991) also display cytotoxic activity *in vivo.* Both ECP and Onconase (Newton *et al.*, 1994) selectively kill Purkinje cells of the cerebellum. In animal models the neurotoxicity is expressed as muscle rigidity, incoordination, ataxia and spasmodic paralysis that are consistent with cerebellar deficit. Neuropathological studies show Purkinje cells are selectively sensitive to these RNases (Gordon phenomenon; Durack *et al.*, 1981).Other homologous enzymes, ribonuclease A (RNase A) and angiogenin, are not cytotoxic. In order to extend our understanding of the

---

† A version of this chapter has been submitted for publication. S.C. Mosimann, D.L. Newton, R.J. Youle & M.N.G. James (1995) X-Ray Crystallographic Structure of Recombinant Eosinophil-Derived Neurotoxin at 1.83Å Resolution. *Biochemistry*

structure and function of cytotoxic, vertebrate RNases, rEDN has been crystallized and its structure has been determined using molecular replacement techniques. The rEDN structure has been refined at 1.83Å resolution to a conventional R-factor ($=\Sigma||F_o|-|F_c||/\Sigma|F_o|$, where $|F_o|$ and $|F_c|$ are the observed and calculated structure factor amplitudes, respectively) of 0.152 for all data with $1.83Å \leq d_{min} \leq 20Å$ and intensities $I \geq \sigma(I)$. The final molecular model of the crystallographic asymmetric unit contains an N-terminal formyl-methionine residue (^{f}Met), all 134 amino acids of EDN, two non-covalently bound sulfate anions and 121 ordered solvent molecules. The molecular conformation of rEDN and its interactions with the sulfate anion ligands are described. We also compare the conformation of rEDN to those of the cytotoxic Onconase (Mosimann *et al.*, 1994) and the non-cytotoxic RNase A (Zegers *et al.*, 1994).

METHODS

The rEDN gene has been synthesized and the rEDN protein expressed as inclusion bodies in *Escherichia coli* BL21. The inclusion bodies have been denatured, renatured and dialyzed before a two step chromatographic purification protocol (D. Newton *et al.*, 1994). The recombinant enzyme is enzymatically active and does not have the post-translational modifications that are present in EDN isolated from human body fluids. Wild-type EDN is glycosylated at C2 of the indole of Trp7 (Hofsteenge *et al.*, 1995) and at five of the six Asn-X-Ser/Thr sequences (where X represents any amino acid except proline); Asn17, Asn59, Asn65, Asn74 and Asn85 are N-glycosylated whereas Asn92 is not glycosylated (Beintema *et al.*, 1988). The recombinant enzyme retains the cytotoxic activities of wild-type EDN.

Crystallization trials were performed using the vapor diffusion technique and hanging drops. A sparse-matrix screening for appropriate crystallization conditions (Jancarik & Kim, 1991) yielded crystals in $(NH_4)_2\ SO_4$ solutions near neutral pH. The crystal used for the high resolution data collection was grown over a 1 mL reservoir

solution containing 40% $(NH_4)_2,SO_4$ 0.1M sodium cacodylate at pH 6.5 and 5% ethanol. The hanging drop, containing initially 3 μL of 21 mg/mL rEDN and 3 μL of the reservoir solution, produced a rod-shaped crystal with dimensions of 0.60 x 0.15 x 0.15 mm. The rEDN crystals are orthorhombic (space group $P2_12_12_1$) and the unit cell dimensions determined from the high-resolution data set are: a=53.42(0.07)Å, b=57.35(0.08)Å, c=42.26(0.06)Å and $\alpha=\beta=\gamma=90°$. The calculated solvent content is 40% and the specific volume (V_m) is 2.04Å^3/Da (Matthews, 1968).

The X-ray intensity data were collected locally on twin San Diego Multiwire Systems (SDMS) multiwire proportional counters. Graphite monochromated $CuK_{\bar{\alpha}}$ X-rays were generated with a Rigaku Rotoflex RU-200BH rotating anode source operated at 40 kV and 150 mA. The data were collected and processed to 1.83Å using the SDMS software (Hamlin, 1985). Statistics for the processing of the native data set is presented in Table 5.1. In the 1.97Å to 1.83Å resolution range, 69% of the unique reflections have $I \geq \sigma(I)$.

The structure of rEDN was determined by the molecular replacement method using the AMORE suite of programs (Navaza *et al.*, 1994). The rotation and translation function solutions were obtained using a truncated model of RNase A. In the search model, residues 17 to 26 were deleted, residues 2, 7, 10, 29, 31, 76, 79 and 98 were truncated to alanine and residues 13, 33, 37, 61, 66, 85 and 90 were modeled as serine. The deleted residues of the RNase search model correspond to an insertion in the RNase A sequence relative to the EDN primary sequence. The residues that were modeled as alanine or serine have large, solvent exposed side chains that are not conserved in the alignment of the sequences of the two enzymes. With this search model, the correct rotation function solution was the fourth largest peak when using data extending from 12 to 2.5Å resolution and an integration radius of 24Å. The correct translation function solution has the highest correlation coefficient (0.386) and the lowest R-factor (52.4). Its correctness was evaluated using packing functions and by calculating and interpreting electron density

Table 5.1

Summary of data collection statistics for the rEDN native data set[+]

	unit cell (Å)			Resol.	No.	Unique	%		
	a	*b*	*c*	(Å)	Refln.	Refln.	Complete	I/σ(I)	R_m[‡]
Native	53.42	57.35	42.26	1.83	58062	10840	90.8	12.3	0.053

[+] Space Group $P2_12_12_1$; wavelength ($CuK_{\tilde{\alpha}}$) = 1.5418Å

[‡] $R_m = \Sigma_{hkl}(\Sigma| I_{obs} - \langle I\rangle|/\Sigma\langle I\rangle)$, where I_{obs} and $\langle I\rangle$ are the observed intensity of a single reflection and the average intensity of all measurements of a reflection and its Freidel mate, respectively.

maps.

The TOM Alberta/Caltech 2.8.0 interactive graphics program (based on Jones, 1985) was utilized to fit the molecular replacement search model to the electron density. The ^{f}Met and Gln116 to Pro122 amino acids were the only residues that could not be fit unambiguously into the initial electron density maps. Data between 8.0 and 1.83Å resolution with intensities greater than 2σ were selected for the initial stages of refinement using the simulated annealing protocol described in the X-PLOR manual (Brünger *et al.*, 1987). During this procedure the crystallographic R-factor ($R=\Sigma||F_o|-|F_c||/\Sigma|F_o|$, where $|F_o|$ and $|F_c|$ are the observed and calculated structure factor amplitudes) was reduced from 0.519 to 0.263. Two sulfate anions and the segment of residues Gln116 to Pro122 were fit into a $3|F_o|-2|F_c|$ electron density map calculated after the molecular dynamics refinement. All subsequent steps in the refinement of rEDN made use of the TNT program suite (Tronrud, 1992) and data in the range $1.83Å \leq d_{min} \leq 20Å$ with intensities greater than 1σ. Interactive graphics refitting sessions utilized electron density maps calculated with $3|F_o|-2|F_c|$, $2|F_o|-|F_c|$ and $|F_o|-|F_c|$ coefficients obtained from TNT.

At several points during the refinement, standard omit maps were calculated (Bhat *et al.*, 1988). In these maps selected atoms were omitted from a short refinement (5 cycles) and the subsequent map calculation. The resulting electron density in the region of the omitted atoms represents a relatively unbiased view of the true electron density at these positions. The residues Ser64 to Lys66, Pro88 to Ile93 and the sulfate anions were all checked against their density in these maps at several stages of the refinement. The side chains of Asn25, Asn39 and Asn104 were also examined in similar omit maps and subsequently modeled as having alternate conformations in the rEDN structure. At occupancies of 0.3 and 0.7 for the two alternate conformations, the average side chain B-factors of each of the conformations are between $17Å^2$ and $26Å^2$. The N-terminal ^{f}Met residue was built into $|F_o|-|F_c|$ difference electron density towards the end of the refinement. Throughout the refinement no effort was made to refine water molecules with alternate

positions and/or variable occupancies. The final model contains 121 ordered, full-occupancy water molecules and has an R-factor of 0.152 for all data with $I/\sigma(I) \geq 1.0$ from 20 to 1.83Å resolution. The stereochemical parameters associated with the final cycle of least-squares refinement are given in Table 5.2. The rEDN coordinates will be deposited in the Protein Data Bank at the Brookhaven National Laboratory (Bernstein *et al.*, 1977).

The least-squares superpositions of RNase A, Onconase and rEDN were computed (Kabsch, 1978) using the O program suite (Jones *et al.*, 1991). The active sites of rEDN and RNase A were superposed using the main-chain and C^{β} atoms of the Thr6 to His15, Arg36 to Thr42 and Val126 to Ile133 residues of rEDN and the equivalent atoms of RNase A (Zegers *et al.*, 1994).

RESULTS

Quality of the refined structure

The final $(3|F_o|-2|F_c|)\ \alpha_{calc}$ electron density is continuous and well-defined for all main-chain atoms of residues 1 to 89 and 92 to 134 at the 1σ contour level. The N-terminal formyl-methionine (^{f}Met) and residues Pro90 and Gln91 have discontinuous and relatively weak electron density. All side-chain atoms of residues in the hydrophobic core and nearly all side-chain atoms of residues on the surface of rEDN have continuous electron density at the same contour level. However, the solvent-exposed side chains of Gln40, Lys66, Gln91 and Arg117 are disordered beyond C^{γ} and have no associated electron density at the 0.9σ contour level. The two non-covalently bound sulfate anions (SO_4(A) and SO_4(B)) have distinct lobes of electron density for their respective oxygen atoms. Figure 5.1a shows the $(|F_o|-|F_c|)\ \alpha_{calc}$ difference electron density contoured at 2.5σ and 8σ for both sulfate anions after refinement with the atoms of the sulfate anions omitted from the structure factor calculations. The refined sulfate anion positions are also displayed in

Table 5.2

Stereochemical statistics for the final cycle of restrained least-squares refinement

R-factor	0.152 for 10627 reflections > 1.0σ|I| between 20.0 - 1.83Å		
resolution			
Model			B-factor (mean)
	541 main-chain atoms		12.7 Å^2
	548 side-chain atoms		17.4 Å^2
	121 waters		38.2 Å^2
	2 sulfate anion		42.7 Å^2
Stereochemistry		rmsd‡	Restraint§
	1 - 2 bond distance	0.014Å	0.70
	Bond Angles	1.87°	0.55
	Pseudorotation Angles	17.11°	0.50
	Trigonal planar groups	0.013Å	1.50
	General planar groups	0.009Å	2.50
	non-bonded contacts	0.019Å	6.00
	Temperature factor	3.04Å^2	0.15

‡rmsd = $\sqrt{\Sigma(X_x - \bar{x})^2 / m}$; the root mean square deviation from expected values in small molecule structures

§Restraint is the relative weight assigned to each stereochemical parameter during the final cycle of refinement.

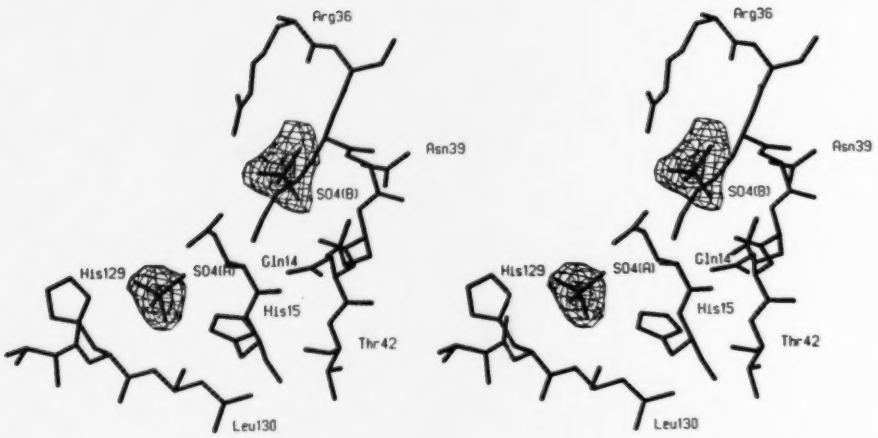


Figure 5.1a The $(|F_o|-|F_c|)\alpha_{calc}$ electron density associated with the sulfate anions after a short refinement and map calculation with the contribution of the sulfate anions omitted. The density is contoured at 2.5σ and 8σ. The figure includes a stick model of the refined rEDN active site and sulfate anions. The are distinct lobes of electron density corresponding to the oxygen atom positions for each sulfate anion. The side chain conformation of Asn39 is one of the two alternate conformations for this side chain.

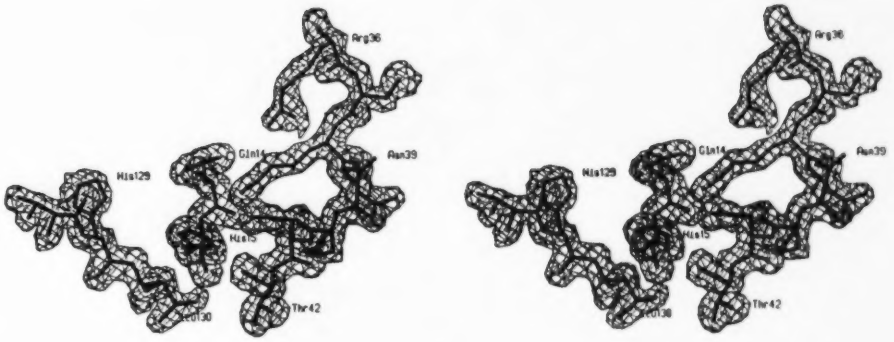


Figure 5.1b The $(2|F_o|-|F_c|)\alpha_{calc}$ electron density associated with the active site residues of rEDN. The active site residues are shown as thick lines and the density is contoured at 1σ. The two alternate conformations of the Asn39 side chain are shown. The Gln40 side chain is partially disordered and does not have associated electron density beyond C^{γ}.

Figure 5.1a superposed on the electron density. Fig5.1b shows the $(2|F_o|-|F_c|)\,\alpha_{calc}$ electron density associated with the active site residues contoured at 1σ.

Figure 5.2 is a plot of the average B-factor of main-chain atoms ($<B_{mc}>$; ($B = 8\pi^2\overline{u^2}$), where $\overline{u^2}$ is the mean-square vibrational amplitude) and average side-chain B-factor ($<B_{sc}>$) versus amino acid sequence number. In both plots, the residues Ser64 to Lys66 and Thr87 to Ile93 possess the largest average B-factors in the structure. These residues are part of two large, solvent-exposed omega loops (Gly56-Lys69, Thr88-Cys96) and include two of the residues with partially disordered side chains, Lys66 and Glu91. The remaining residues with $<B_{mc}> \geq 20\,Å^2$ are Gly75, Ser76, Glu116 to Arg118, the N-terminal ^{f}Met and the C-terminal Ile133, Ile134. Residues Gly75 and Ser76 are part of a loop structure and residues Glu116 to Arg118 are part of a helical turn.

Figure 5.3 shows the Ramachandran plot for the 118 non-glycine, non-proline residues of rEDN. The main-chain φ,ψ dihedral angles populate the core (103 of 118 residues) and allowed (15 of 118 residues) regions of the Ramachandran plot as defined in PROCHECK (Laskowski *et al.*, 1993). The EDN primary sequence contains 2 glycine and 12 proline residues. The NH of Gly56 is 2.78Å from Asn53O and Gly56 adopts a 3_{10}-helical conformation (φ = -84°, ψ = -9°) at the end of helix α2 (residues Thr47-Cys55). Gly75 has a conformation (φ = 84°, ψ = -49°) not normally allowed to non-glycine residues and Gly75O forms a single intramolecular hydrogen bond with His73N$^{\varepsilon 2}$ (3.06Å). The 12 proline residues each adopt *trans*-peptide conformations and have φ, ψ dihedral angles typical of proline residues. Eight of the proline residues are found in loop structures and four in segments of β-strand. Pro102, Pro124 and Pro127 are parts of solvent accessible β-strands on the edge of their β-sheet. Pro79 is the second residue of a central β-strand and is also solvent exposed.

Molecular conformation of rEDN

The polypeptide fold of rEDN is shown in Figures 5.4 and an all atom stereo

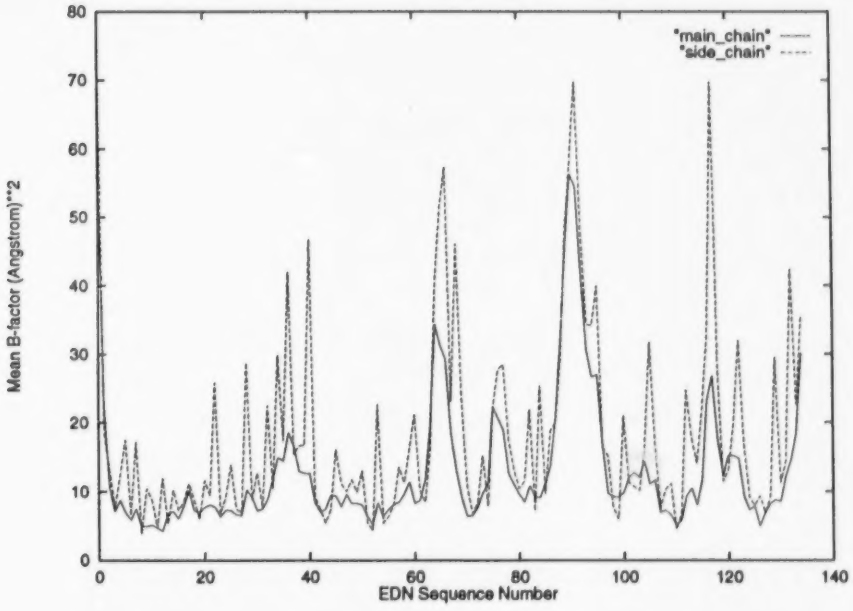


Figure 5.2 The average main-chain B-factor (solid line) and average side-chain B-factor (dashed line) versus sequence number plots for the rEDN structure. The segments Ser64 to Lys66 and Thr87 to Ile93 have the largest average main-chain B-factors and are parts of two long loop structures. The side chains of Lys66, Gln92 and Arg117 are disordered beyond C^{γ} and represent the largest average side-chain B-factors.

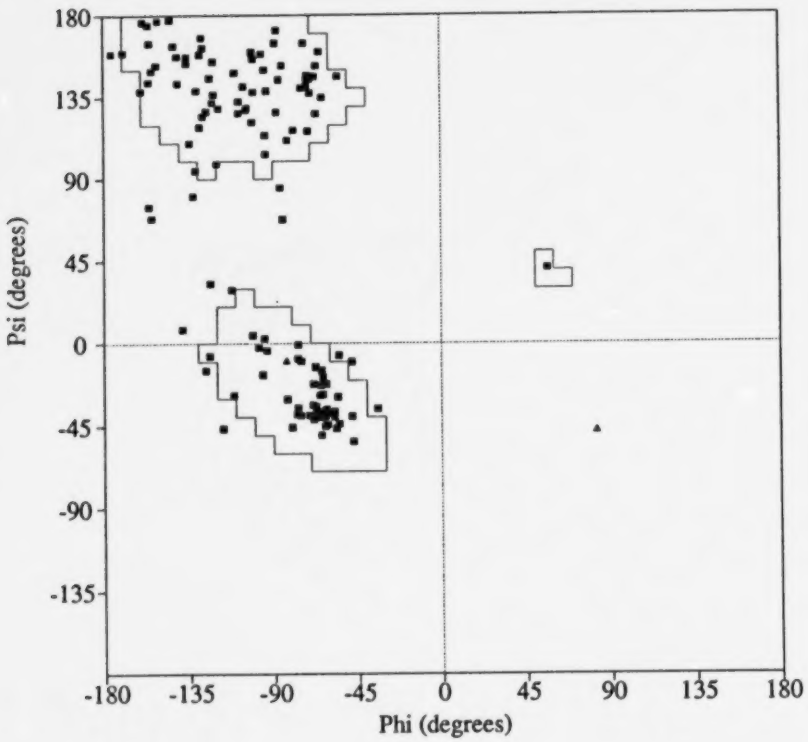


Figure 5.3 A Ramachandran plot of the ϕ, ψ dihedral angles of rEDN. The two glycine residues of rEDN are indicated by filled triangles and the remaining residues of rEDN are indicated by filled squares. The contours represent the most favorable regions of the plot as defined in PROCHECK (Laskowski *et al.*, 1993) and include 103 of the 118 non-glycine, non-proline residues in rEDN. The remaining 15 residues are in the "allowed" regions of the plot (contour not shown).

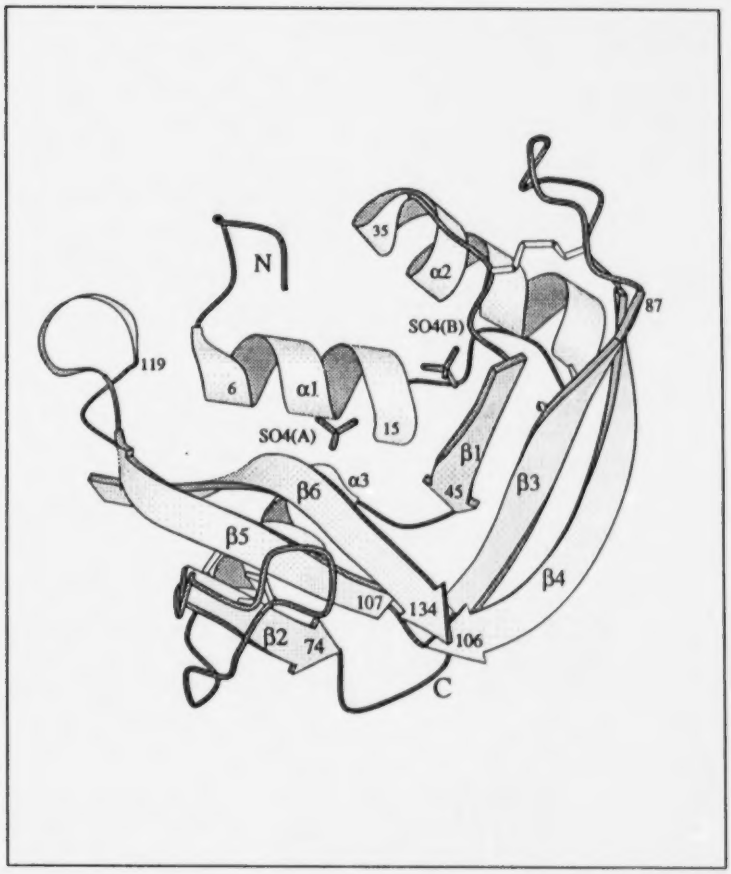


Figure 5.4 A molscript (Kraulis, 1991) diagram of the rEDN polypeptide fold. The helices, strands and loop structures of rEDN are shown as helical ribbons, arrows and filled lines, respectively. The four disulfide bridges and the two non-covalently bound sulfate anions are represented as rods. The helical turn Asp115 to Asp119 (shown as a one turn helix) is part of a nine residue insertion that is unique to EDN.

representation is presented in Figure 5.5. Table 5.3 lists the secondary structural elements present in the rEDN structure as defined by PROCHECK and Table 5.4 lists the β-turns and identifies omega loops (Sibanda & Thronton, 1985; Leszczynski & Rose, 1986). rEDN is comprised of an N-terminal α-

helix and two related lobes. Each lobe contains a three-stranded antiparallel β-sheet and an α-helix with a conserved disulfide bridge between the central strand of the β-sheet and the α-helix. The strands β1, β3, β4 and helix α2 form lobe 1 and strands β2, β5, β6 and helix α3 comprise lobe 2. The conserved disulfide bridges of the two lobes are Cys23-Cys83 (lobe 1) and Cys55-Cys111 (lobe 2) and they share left-handed χ_3 torsion angles of -81° and -78°, respectively. Helix α1 packs predominantly against lobe 2 where it contributes residues to the hydrophobic core. Residues at the C-terminus of helix α1 (Gln14, His15, Ile16) pack against amino acids of lobe 1. The active site of rEDN is located at the junction of lobe 1, lobe 2 and the N-terminal α-helix where a cleft is formed running parallel to strand β6.

The remaining cysteine residues of rEDN form two additional disulfide bonds that help define the tertiary structure. The disulfide Cys37-Cys96 is between the loop following helix α2 and the central strand (β3) of the β-sheet of lobe 1. The disulfide Cys62-Cys71 is part of the large loop (Gly56-Lys69) between helix α3 and strand β2, a strand at the edge of the β-sheet of lobe 2. These two disulfide bridges have right-handed χ_3 torsion angles of 88° and 104°, respectively .

The stretch of polar residues from Asp115 to Tyr123 (Asp-Gln-Arg-Arg-Asp-Pro-Pro-Gln-Tyr) forms a single helical-turn followed by a type III β-turn like structure (Figure 5.6). This compact folding unit corresponds to an omega loop. Asp115O forms a hydrogen bond with Asp119N (2.89 Å) and Asp115$O^{\delta 1}$ hydrogen bonds to Arg118N (2.54 Å) initiating helical turn. Five additional polar interactions stabilize the conformation of this loop: Arg114$N^{\eta 1}$ donates a hydrogen bond to Gln116O (2.82 Å); Arg118$^{\eta 1}$ is 2.97 Å from Pro3O; Asp119$O^{\delta 2}$ accepts a hydrogen bond from Asp115N (2.84 Å) and

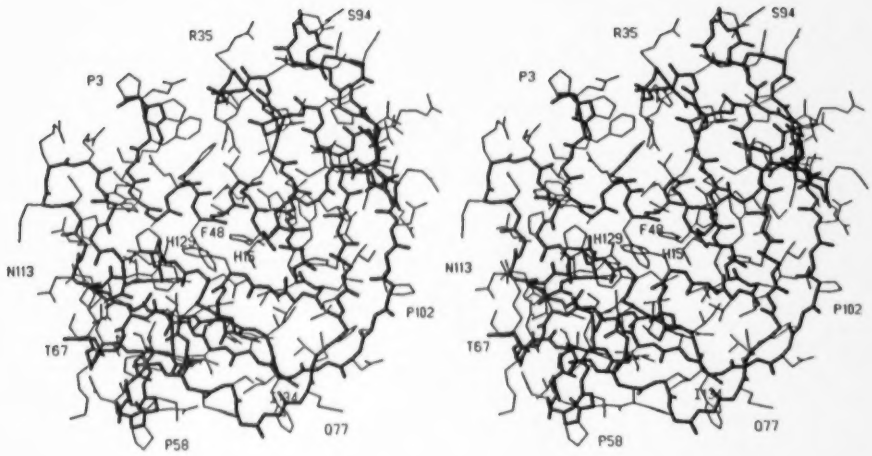


Figure 5.5 A stereo diagram of the all-atom structure of rEDN. The bonds between the main-chain atoms are shown as thick lines; those connecting the side-chain atoms are thin lines. The view is the same as that in Figure 5.4. The active site His15 and His129 and other selected residues are indicated.

Table 5.3

Secondary structures of rEDN

Helices		Strands		loop structures	
α1	Thr6-Ile16	β1	Asn39-Leu45	l 1	Lys1-Phe5
α2	Gln22-Arg35	β2	Asn70-Asn74	l 2	Asn17-Gln21
α3	Thr47-Cys55	β3	Val78-Thr87	l 3	Arg36-Lys38
		β4	Tyr97-Met104	l 4	Gly56-Lys69‡†
		β5	Phe105-Arg114	l 5	Gly75-Gln77
		β6	Pro124-Ile134	l 6	Pro88-Cys96‡†
				l 7	Asp115-Tyr123§†

‡ Residues Pro2 to Phe5, Asn62 to Asn65, Asn65 to Arg68 and Gln92 to Asn95 are type I β-turns.

§ Residues Asp115 to Asp119 form a helical turn and Pro120 to Tyr123 form a type III β-turn.

† Loops l 4, l 6, and l 7 are omega loops (Leszczynski & Rose, 1986).

Table 5.4

β-turn structures of rEDN

Type	Residues *i, i+1, i+2, i+3*	ϕ_{i+1}, ψ_{i+1}	ϕ_{i+2}, ψ_{i+2}
I	Pro2 Pro3 Gln4 Phe5	-66°, -21°	-79°, -7°
I	Cys62 Pro63 Ser64 Asn65	-54°, -35°	-85°, -10°
I	Asn65 Lys66 Thr67 Arg68	-46°, -11°	-98°, -1°
I	Asn92 Ile93 Ser94 Asn95	-42°, -33°	-66°, -12°
III	Pro120 Pro121 Gln122 Tyr123	-59°, -26°	-44°, -43°

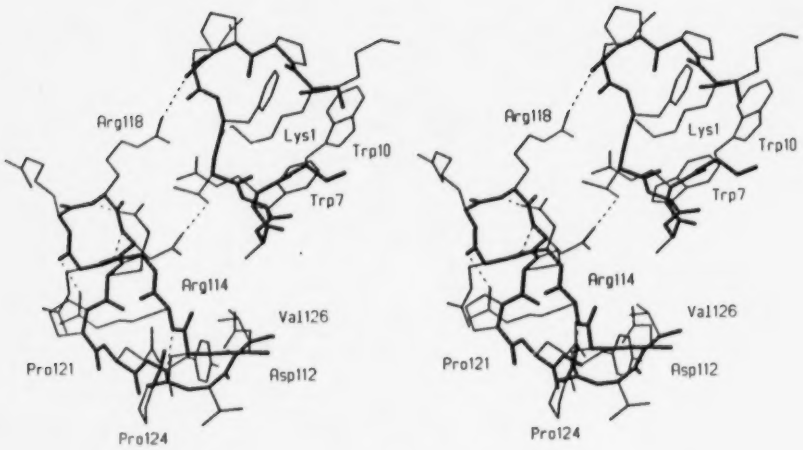


Figure 5.6 A stereo diagram of the residues of Asp115 to Tyr123 of rEDN and their packing interaction with the N-terminus. The main-chain and side-chain bonds are represented by thick and thin lines, respectively and hydrogen bonds are indicated as dashed lines. The residues of Asp115 to Asp119 form a single helical turn and Pro120 to Tyr123 forms a type III β-turn. The side chain of Asp119 forms a hydrogen bond to the side chain of Thr6 and the negatively charged carboxylate balances the positively charged dipole moment of the helix α1.

Asp119O$^{\delta 1}$ is positioned opposite the dipole of helix α1 where it also forms a hydrogen bond with Thr6O$^{\gamma 1}$ (2.68 Å). The segment Pro120 to Tyr123 breaks the helix and adopts a type III turn like conformation that changes the direction of the polypeptide chain and leads into strand β6.

The loop 17, and in particular the helical turn, packs against and influences the conformation of N-terminus of EDN. The side chains of Asp115, Arg118 and Asp119 pack against the N-terminus of helix α1. Helix α1 is initiated by helix-capping hydrogen bonds from the Thr6 O$^{\gamma 1}$ and O atoms to Gln9N and Trp10N, respectively. The conformation of Thr6 is further fixed by the Thr6N to Gln9O$^{\varepsilon 1}$ hydrogen bond. Residues Pro2 to Phe5 adopt a type I β-turn conformation and the Lys1N$^{\zeta}$ forms a hydrogen bond with Phe5O. The conformation of the N-terminus is also stabilized by aromatic stacking interactions between the Phe5 and Trp10 rings (Burley & Petsko, 1986). The side chain of ^{f}Met does not interact with residues of the N-terminus or the residues Asp115 to Tyr123.

The omega loop, 14 of rEDN contains a disulfide bridge (Cys62-Cys71) and consecutive type I β-turns. In loop 14, Asn65N and Asn65O are 3.1Å and 2.8Å from Cys62O and Arg68N, respectively. Another omega loop, 16 contains Cys96 of the conserved Cys37-Cys96 disulfide and a single type I β-turn. The O atom of Asn92 of loop 16 is 2.5Å from Asn95N. The only other non-helical, mainchain *i* to *i+3* hydrogen bond in the rEDN structure is between Ile93O and Cys96N (3.0Å).

Sulfate anion binding sites

Figure 5.7 shows the hydrogen bonding network between rEDN and the two, non-covalently bound sulfate anions (SO_4(A) and SO_4(B)). Table 5.5 presents a list of the hydrogen bonds to the anions. The anions are presumed to be sulfate based upon the presence of $(NH_4)_2SO_4$ (and the absence of phosphate) in the crystallization media. The SO_4(A) anion binds at the active site of rEDN where Gln14N$^{\varepsilon 2}$ and His15N$^{\varepsilon 2}$ of helix α1 are 3.37 Å and 3.08 Å from O4 and O1 of the SO_4(A) anion,

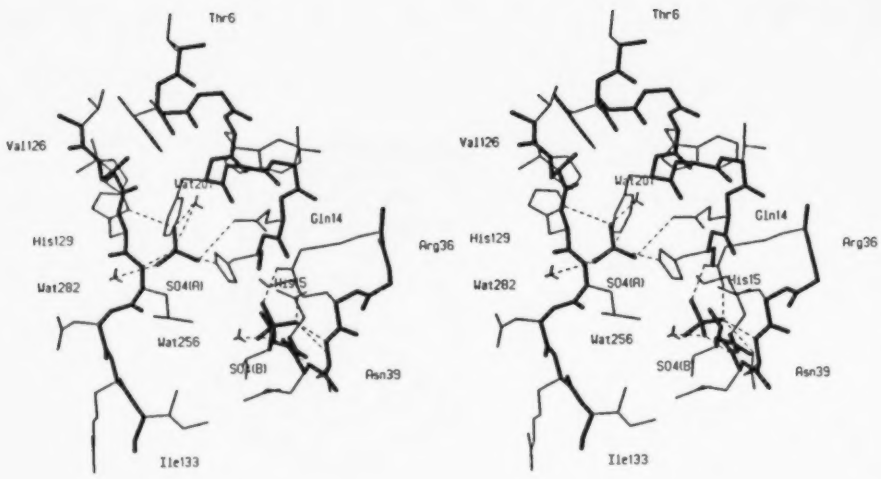


Figure 5.7 The interactions between rEDN and the non-covalently bound SO_4(A) and SO_4(B) anions are shown in a stereo diagram. The thick lines represent the main chain of rEDN the thin lines represent the side chains of rEDN and its anions and hydrogen bonds are represented as dashed lines. The SO_4(A) anion at the active site contacts the side chains of Gln14, His15 and His129 and the main chain of Leu130. The anion SO_4(B) is bound by the main chain of Asn39 and Gln40 and the side-chain atoms of Arg36 and Asn39. Ordered solvent molecules that make contacts with either anion are shown as small tetrahedra.

Table 5.5

The sulfate anion contacts distances (Å) in EDN

SO_4(A)		SO_4(B)	
O1		O1	
His15N$^{\varepsilon 2}$	3.08	Arg36N$^{\eta 2}$	2.50
Leu130N	2.61		
Wat201	3.13		
O2		O2	
Wat282	2.73	Gln40N	3.12
His129N$^{\delta 1}$	3.33	Wat256	3.06
O3		O3	
His129N$^{\delta 1}$	2.63	†Asn39N$^{\delta 2}$	3.08
Wat201	2.96		
O4		O4	
Gln14N$^{\varepsilon 2}$	3.23	Arg36N$^{\eta 1}$	2.97
His15N$^{\varepsilon 2}$	3.47	Asn39N	2.60
		†Asn39N$^{\delta 2}$	2.69

† The Asn39 side chain is modeled as two separate conformations in the rEDN structure. The Asn39 side chain conformation that contacts the SO_4(B) is the minor conformer (occupancy=0.3).

respectively. The remaining two hydrogen bonds between rEDN and SO_4(A) are from His129 $N^{\delta 1}$ (2.63Å) and Leu130 N (2.61 Å) of strand β6 to O3 and O1 of the anion, respectively. Two ordered solvent molecules, Wat201 and Wat282 also contact the bound sulfate anion. Wat201 accepts a 3.05Å hydrogen bond from Gln14 $N^{\varepsilon 2}$ and is ≤3.30Å from O1 and O3 of the anion. This water has counterparts in both the RNase A and the Onconase structures. Wat282 forms a 2.73Å hydrogen bond with O2. The side chain of the catalytically important Lys38 is well-defined in the electron density maps and the N^{ζ} atom forms a hydrogen bond with the Gln14 $O^{\varepsilon 1}$. In this conformation, Lys38 N^{ζ} is 3.9 Å from O4 of the anion.

The anion SO_4(B) is bound to the surface of a single rEDN molecule by three main-chain hydrogen bonds and the side-chain groups of Arg36 and Asn39. Gln40N is 3.12Å from O2 and Asn39N is 2.60Å and 3.28Å from O4 and O2, respectively. Arg36 $N^{\eta 1}$ and Arg36 $N^{\eta 2}$ are within hydrogen bond distance of O4 and O1, respectively. The alternate side-chain conformation (occupancy = 0.30) of Asn39 provides two additional contact with O3 and O4. Wat256 is the only ordered solvent molecule hydrogen bonding with SO_4(B). The distance between the two sulfur atoms of SO_4(A) and SO_4(B) is 9.4Å. The second anion binding site has not been observed in structures of other members of the RNase superfamily.

Intermolecular packing in rEDN

Figure 5.8 is a stereo diagram representing the intermolecular packing within the rEDN crystal. Table 5.6 lists the polar hydrogen-bonding interactions and van der Waals contacts between symmetry-related molecules. Two hydrogen bonds are accepted by Gln12 $O^{\varepsilon 1}$ (Arg97 N^{ε}, Arg97 $N^{\eta 2}$) and by Arg35O (Arg114 N^{ε}, Gln116 $N^{\varepsilon 2}$), respectively and Gln21 $N^{\varepsilon 2}$ donates hydrogen bonds to Cys55O Asn57O. Thr61 also forms a pair of hydrogen bonds and is involved in the sole polar intermolecular contact between main-chain atoms (Thr61N to Thr86O). Gln12, Gln21, Arg35 and Thr61 account for 8 of the

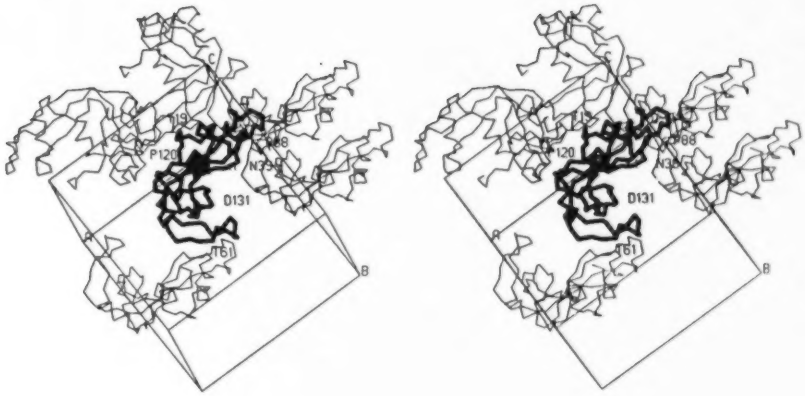


Figure 5.8 A C^{α} backbone trace representing the intermolecular packing in the rEDN crystal is shown as a stereo diagram. The rEDN molecule at x, y, z is shown in thick lines and the five symmetry related molecules that make contacts with it are represented as thin lines. Residues in the rEDN molecule in thick lines are labeled as are the unit cell axes.

Table 5.6

The intermolecular contacts in EDN

Polar <3.30 Å		van der Waals < 3.90 Å	
-x+1/2, -y, z+1/2			
Thr13 $O^{\gamma 1}$ - Gln100 $N^{\varepsilon 2}$	3.29	Phe5 $C^{\delta 2}$ - Gln22 $N^{\varepsilon 2}$	3.11
Gln12 $O^{\varepsilon 1}$ - Arg97 $N^{\eta 2}$	3.13	Asn17 $N^{\delta 2}$ - Ala99 C^{α}	3.74
Gln12 $O^{\varepsilon 1}$ - Arg97 N^{ε}	2.87	Val29 $C^{\gamma 1}$ - Gln100 O	3.40
		Tyr33 C^{ζ} - Pro102 C^{γ}	3.84
		Tyr33 OH - Gln100 C^{γ}	3.41
		Phe48 $C^{\delta 2}$ - Lys66 O	3.52
-x+1, y-1/2, -z+1/2			
Ile16 O - Lys69 N^{ζ}	3.29	Ser19 O - Pro58 C^{γ}	3.31
Gln21 $N^{\varepsilon 2}$ - Asn57 O	2.69	Thr47 $C^{\gamma 2}$ -Lys66 O	3.65
Gln21 $N^{\varepsilon 2}$ - Cys55 O	3.15		
Thr46 O - Lys69 N^{ζ}	3.27		
Asn50 $O^{\delta 1}$ - Thr67 O	2.96		
x-1/2, -y+1/2, -z+1			
Arg35 O - Gln116 $N^{\varepsilon 2}$	3.19	Pro3 C - Gly56 C^{α}	3.80
Arg35 O - Arg114 $N^{\eta 2}$	3.06	Pro3 C^{β} - Val52 O	3.32
		Gln4 C^{γ} - Gly56 O	3.70
		Tyr33 O - Pro124 C^{β}	3.45
		Arg35 C^{γ} - Pro124 O	3.61
x+1/2, -y+1/2, -z			
Thr61 N - Thr86 O	2.92	Met60 C^{β} - Thr86 O	3.44
Thr61 $O^{\gamma 1}$ - Thr86 $O^{\gamma 1}$	3.16		

12 unique intermolecular hydrogen bonds and each residue is in contact with a different symmetry related molecule. The Pro3 and Gln4 residues at the N-terminus and Tyr33 account for 6 of the 14 van der Waals contacts <3.90 Å. Four additional proline residues Pro58, Pro88, Pro102 and Pro124 also participate in packing interactions. The shortest intermolecular van der Waals contact is 3.11Å, Phe5$C^{\delta 2}$ to Gln22$N^{\varepsilon 2}$.

Superposition of rEDN with RNase A and Onconase

Figure 5.9 is a stereo diagram of the least-squares superposed X-ray crystallographic structures of rEDN, RNase A and Onconase. The superpositions of rEDN with RNase A and Onconase have rmsds of 1.27 Å and 1.42 Å for the 106 (of 118) and 86 (of 101) common C^{α} atoms that can be superposed within 3.8 Å, respectively. The residues in helices α1 to α3 and those in strands β1 to β6 account for nearly all the equivalent residues in the three structures. The residues Ile30 to Arg35 of rEDN extend helix α2 relative to RNase A and are the only residues of these secondary structures that do not superpose with their counterparts in RNase A and Onconase. The loops l1, l2, l4, l6 and l7 (Table 5.3) are the sites of insertions and deletions in the three structures and include those remaining residues that do not superpose within 3.8 Å in the overlaps. The conformation and intramolecular packing between the N-terminus (loop l1) and the 9 residue insertion of (helix α4 and loop l7) rEDN have been described. The loop l2 of rEDN is 6 residues shorter and 1 residue shorter than the equivalent loops in RNase A and Onconase, respectively; the conformation of l2 is completely different in all three structures (Figure 5.9). The relatively large loops, l4 and l6 of rEDN are 2 residues longer than the equivalent loops in RNase A. In each case, the additional residues of rEDN result in conformational differences along the entire length of the loop. The loops l4 and l6 of Onconase are 8 residues shorter and 6 residues shorter than the equivalent loops of rEDN and have very different conformations in RNase A and Onconase (Mosimann *et al.*, 1995).

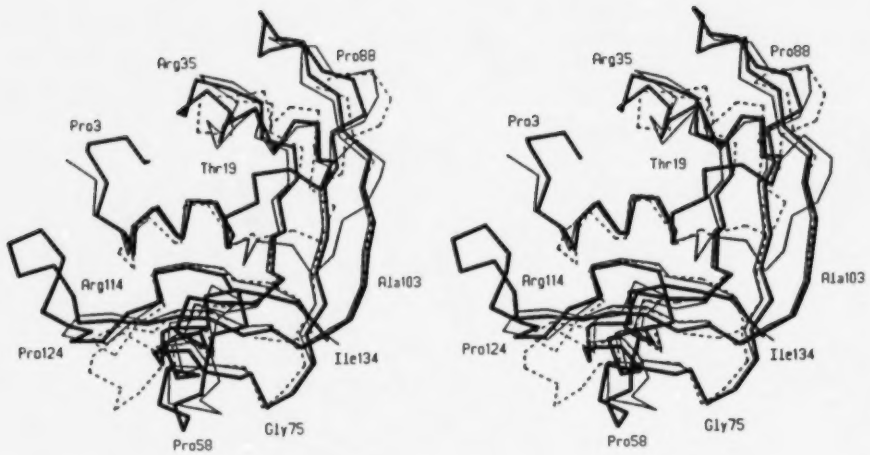


Figure 5.9 The least-squares superposition of rEDN, RNase A (Zegers *et al.*, 1994) and Onconase (Mosimann *et al.*, 1994) based on their C^{α} coordinates. rEDN, RNase A and Onconase are represented by thick, thin and dashed lines, respectively. The α-helices and β-strands of the structures superpose relatively closely whereas the main conformational differences are in the connecting loop structures.

Figure 5.10 shows the superposed active sites of RNase A (Zegers *et al.*, 1994) and rEDN. The mainchain and C^{β} atoms of residues Thr6-His15, Arg36-Thr42 and Val126-Ile133 of rEDN were superposed with the equivalent residues of RNase A by least-squares. The amino acid sequence identity for these 25 residues is 64% and the rmsd is 0.60Å for 125 common atoms. The sulfate anion of the RNase A structure and SO_4(A) of rEDN have nearly identical conformations and their sulfur atoms differ in their positions by 0.6Å. This difference is less than the 0.9Å that separates Asp121C^{α} (RNase A) and Asp131C^{α} (rEDN). The side chains of Gln11, His12, His119 and the main chain of Phe120 of RNase A form hydrogen bonds with the anion. Gln14, His15, His129 and Leu130 of rEDN make equivalent contacts with SO_4(A). His119 of RNase A adopts two alternate conformations in the RNase A structure (Zegers *et al.*, 1994). The His129 conformation of rEDN corresponds to the catalytically inactive conformation of His119 of RNase A. His129 can adopt the active conformation observed in RNase A by rotation about its χ_1 and χ_2 torsion angles. In order for His129 to adopt the active conformation, Wat282 would have to be displaced (Figure 5.7).

In the superposition of RNase A and EDN, the SO_4(B) of rEDN is less than 1Å from the side chain of Pro42 of RNase A. Pro42 of RNase A is the residue equivalent to Asn39 of rEDN. In the crystal structure of rEDN, the NH and $N^{\delta 2}$ atoms of Asn39 donate hydrogen bonds to SO_4(B) (Figure 5.7). Obviously, Pro42 of RNase A cannot act as a hydrogen bond donor and as a result, RNase A cannot bind an anion in a manner equivalent to the SO_4(B) binding observed in rEDN.

rEDN has been superposed with the RNase A-thymidylic acid tetramer ($d(Tp)_4$) complex structure (Birdsall *et al.*, 1992) by least squares. Figure 11 is a stereo diagram of rEDN and three and of the four superposed thymidilic acids. The base, ribose and phosphate moieties shown in Figure 11 are B_2, B_1, B_0, R_2, R_1, R_0 and P_1, P_0, P_{-1} using the nomenclature of Richards & Wykoff, 1971 (Figure 12). The conformation of $d(Tp)_4$ has not been altered though one nucleotide has been omitted for clarity.

In the rEDN-d(Tp)$_3$ model, the B_2, B_1, R_2, R_1, P_1 and P_{-1} moieties can make specific interactions with rEDN. The bases both adopt *anti* conformation. The side chains of Asn70 and Asp112 are within the B_2 subsite and the N and $O^{\gamma 1}$ atoms of the invariant Thr42 of rEDN complement the hydrogen bonding potential of the O2 and N3 atoms of B_1. The hydrogen bonding potential of the O4 atom of B_1 is unsatisfied in Figure 10. The R_2 and R_1 moieties are involved in van der Waals contacts with the Trp7 ring and Leu130, respectively. The P_1 and P_{-1} phosphates in the rEDN-d(Tp)$_3$ model have positions similar to the SO_4(A) and SO_4(B) anion positions observed in the rEDN. The P_1 phosphate of the model is located in the rEDN active site and can hydrogen bond with Gln14$N^{\epsilon 2}$, His15$N^{\epsilon 2}$, His129$N^{\delta 1}$ and Leu130N. The P_{-1} phosphorus atom of the model is more than 3.0Å from the sulfur atom position of SO_4(B). In the RNase A-d(Tp)$_4$ structure, the P_{-1} phosphate makes van der Waals contacts with Pro42.

DISCUSSION

The polypeptide fold of rEDN is related to those observed in the homologous RNase A (Zegers *et al.*, 1994), Onconase (Mosimann *et al.*, 1994) and angiogenin (Acharya *et al.*, 1994) X-ray crystallographic structures. These enzymes share less than 35% amino-acid sequence identity and they range in size from 104 residues (Onconase) to 134 residues (EDN). Given the relatively low sequence identity and the size variation it is not surprising that the molecular surface of EDN is different from that of other pyrimidine specific RNases of known structure. The loops l2, l3 and l6 of EDN have significantly different conformations from those observed in structures of other vertebrate RNases. This is significant since enzymes belonging to the vertebrate RNase superfamily have a variety of biological functions and properties in addition to that of digesting RNA. The non-ribonucleolytic properties of several of the vertebrate enzymes are known to be associated with their specific binding to targets on the cell surface (Wu *et al.*, 1993; Hallahan *et al.*,

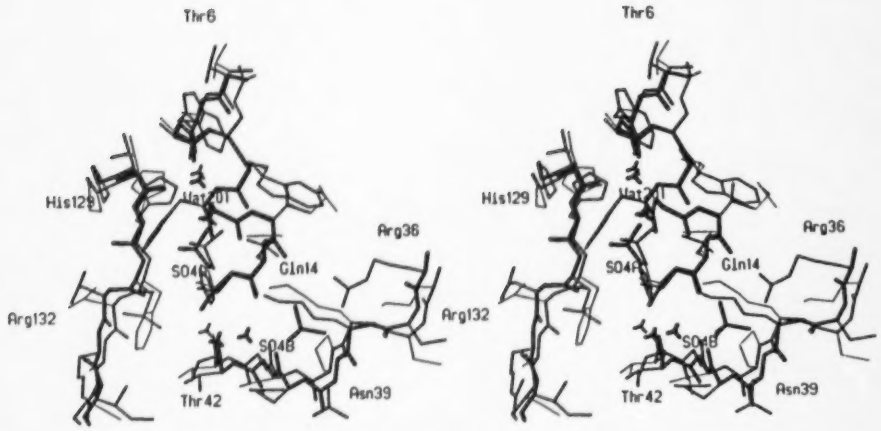


Figure 5.10 A stereo diagram of the least-squares superposed active site residues and bound anions of rEDN and RNase A are shown from the same vantage point as that in Figure 5.7. The rmsd for the 125 main-chain and C^{β} atoms is 0.60Å. The thick lines represent rEDN and its anions and the thin lines represent RNase A and its anion. The conformations of the catalytic residue and the positions of the active site anions and are closely similar in the two structures. The conformation of His129 of rEDN is equivalent to one of the two alternate conformations (inactive) of the His119 (RNase A). The SO_4(B) anion of rEDN makes a number of too-close contacts with Pro42 of RNase A. This proline residue probably prevents RNase A from binding an anion in a manner analogous to the binding of SO_4(B) to rEDN.

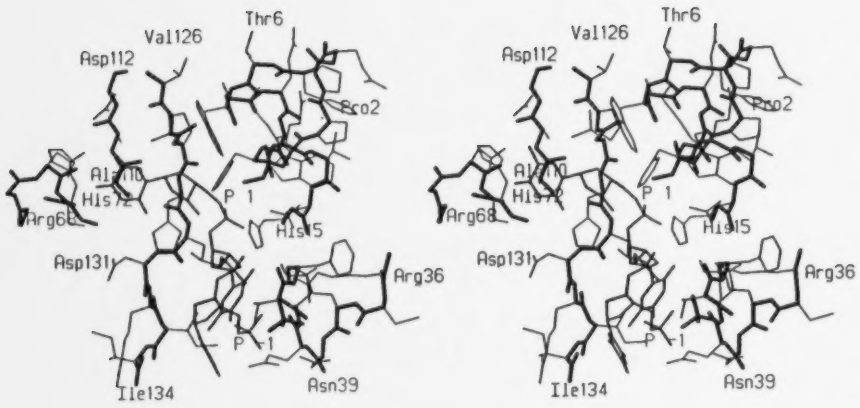
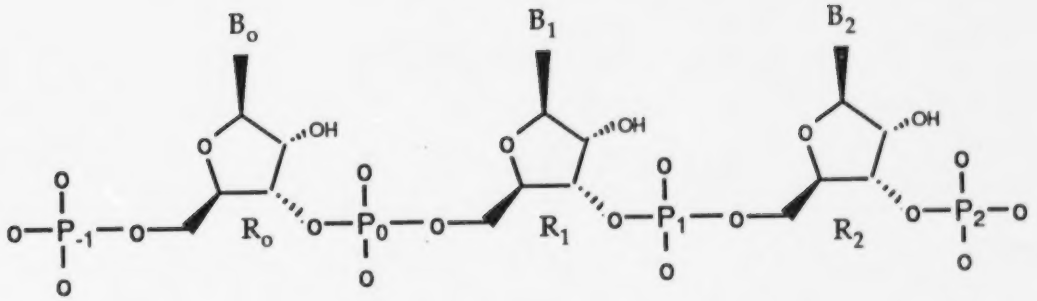


Figure 5.11 A stereo diagram of the rEDN-d(Tp)$_3$ model based upon the RNase A-(Tp)$_4$ structure (Birdsall *et al.*, 1992). The rEDN backbone is shown as thick lines and the thin lines represent the side chain atoms. The d(Tp)$_3$ inhibitor is shown in lines of intermediate thickness and the P_1 and P_{-1} phosphates are labeled. The SO_4(B) anion of rEDN is more than 3Å from the nearest anion, the P_{-1} phosphate. The Asn70 and Asp112 residues of rEDN are within the B_2 subsite and Thr42 and Gln40 are in the B_1 subsite.


B0
B1
B2
OH
OH
OH
O
O
O

1991; Titani *et al.*, 1987). These loops may play a role in the function of EDN. Interestingly, EDN and Onconase share little structural homology other than their general fold yet EDN, ECP and Onconase represent the only neurotoxins within this superfamily. Additionally, the large insertion unique to the non-secretory RNases (eg. EDN: Asp115 to Tyr123) may contribute to their unique properties. At present these residues have no known biological role.

rEDN has two discrete SO_4 anion binding sites. The active site SO_4(A) and the SO_4(B) anions observed in the rEDN structure are wholly bound by a single rEDN molecule and do not make contacts with symmetry related molecules. In the structures of RNase A-oligodeoxynucleotide complexes (Birdsall *et al.*, 1993; Fontecelli-Camps *et al.*, 1994), phosphate, ribose and base subsites have been identified for extended substrates (Figure 11). Phosphate subsites corresponding to P_1 and P_2 have been observed while the P_0 and P_3 phosphate moieties do not bind to specific subsites. The SO_4(A) (P_1) and SO_4(B) anions are 9.4Å apart in rEDN structure and this distance is considerably larger than distance between consecutive phosphates of RNA or DNA when in an extended conformation. This would indicate the SO_4(B) binding site of rEDN is not P_0 or P_2. In the RNase A-d(Tp)$_4$ structure, the P_{-1} phosphate makes van der Waals contacts with the side chain of Pro42. The residue that is structurally equivalent to Pro42 of RNase A is Asn39 and it donates hydrogen bonds to SO_4(B) through its main-chain and side-chain atoms (Figure 5.10). In Figure 11, the SO_4(B) anion of rEDN is nearest to the P_{-1} phosphate of the RNase A-d(Tp)$_4$ model (>3.0 Å between sulfur to phosphate). As a result, the SO_4(B) binding site likely correspond to the P_{-1} subsite of rEDN. Within the superfamily, the residue equivalent to Pro42 of RNase A is not highly conserved. It is present in pancreatic enzymes and angiogenins but not in amphibian or non-secretory RNases. Thus, the P_{-1} phosphate binding in RNase A may not be representative of the P_{-1} phosphate binding in all related enzymes. Instead, relatives that do not have an imino acid at this position may bind the P_{-1} phosphate similar to the observed SO_4(B) binding mode

of rEDN.

The enzymatic activity and specificity of rEDN towards diribonucleotide substrates are comparable to those of RNase A. rEDN has a slight preference for cytidine in the B_1 subsite and prefers adenosine in the B_2 subsite. In the superposition of rEDN and $d(Tp)_4$ (Figure 12) the N2 and O3 atoms of the B_1 base can hydrogen bonded to the invariant Thr42. In RNase A, O4 of B_1 hydrogen bonds to a solvent molecule that hydrogen bonds to Ser123 (Zegers *et al.*, 1994). The equivalent residue of rEDN is Ile133 and it cannot make analogous interactions. The nearest potential protein hydrogen bond donor or acceptor of rEDN is Gln40. This is interesting as Gln40 can adopt conformations that allow it to hydrogen bond to $SO_4(B)$ yet does not. In the X-ray structure of rEDN, the Gln40 side chain is solvent exposed and disordered beyond C^{γ}. Perhaps Gln40 is responsible for binding the O4 of uracil or N4 of cytosine. Like Ser123 of RNase A, Gln40 can either accept or donate hydrogen bonds. This suggests both cytosine and uracil can be accommodated within the B_1 subsite and explains the observed specificity of rEDN (Table 1.4). In RNase A complex structures, the B_2 base hydrogen bonds to Asn71 and Glu111. The equivalent residues of rEDN are Asn70 and Asp112. In Figure 11, Asn70 of rEDN can hydrogen bond to O4 atom while Asp112 is directed at the C5 methyl group. The B_2 base of the RNase A-$d(Tp)_4$ structure is rotated by 180° when compared to the high resolution RNase A-d(CpA) structure. The pseudorotation angle of the sugars R_1 and R_2 and the glycosyl torsion angle of the bases B_1 and B_2 are comparable in these structures. The conformational difference arises from large differences in the β (P-O5'-C5'-C4') and γ torsion angles (O5'-C5'-C4'-C3') of the nucleic acid backbone of these structures. If the B_2 base binds as seen in the high resolution d(CpA) structure, Asp112 will be in close proximity to the N3 atom. Arg68 is an additional rEDN residues that may be near enough to contribute to the binding the B_2 base .

In a recent comparative molecular modeling assessment, the loops 12, 13 and the omega loop 14, 16 and 17 (the large insertion of rEDN) were not successfully modeled by

the participating groups (Mosimann *et al.*, 1995). These large loops are the sites of insertions and deletions in comparisons of the primary sequences of rEDN, RNase A and Onconase and correspond to the regions of the greatest divergence in structural superpositions of these enzymes (Figure 5.9). In rEDN, it appears the loops of RNase A and Onconase represent poor models of the conformations of loops in rEDN as a result of the low sequence identity and the presence of insertions and/or deletions. The conformation of the large insertion of rEDN is exceedingly difficult as it packs against additional inserted residues at the N-terminus (Figure 5.6). A successful prediction of the conformation of this loop would require the correct simultaneous prediction of the conformation of the N-terminus as well. In many ways this is a small version of the folding problem. These difficulties suggest the prediction of large, divergent loop structures is not yet possible using current methods.

REFERENCES

Acharya, K.R., Shapiro, R., Allen, S.C., Riordan, J.F., Vallee, B.L. (1994). Crystal structure of human angiogenin reveals the structural basis for its functional divergence from ribonuclease. *Proc. Natl. Acad. Sci. U.S.A.* **91**:2915-2919.

Bhat, T.N. (1988) Calculation of an omit map. *J. Appl. Crystallog.* **21**, 279-281.

Birdsall, D.L. & McPherson, A. (1992). Crystal structure disposition of thymidylic acid tetramer in complex with ribonuclease A. *J. Biol. Chem.* **267**, 22230-22236.

Brünger, A., Kuriyan, J. & Karplus, M. (1987). Crystallographic R factor refinement by molecular dynamics. *Science* **235**, 458-460.

Beintema, J.J., Schuller, C., Irie, M. & Carsana, A. (1988). Molecular evolution of the ribonuclease superfamily. *Prog. Biophy. Molec. Biol.* **51**, 165-192.

Bernstein, F.C., Koetzle, T.F., Williams, G.J.B., Meyer Jr., E.F., Brice, M.D., Rodgers, J.R., Kennard, O., Shimanouchi, T. and Tasumi, M. (1977). The Protein Data Bank: A computer-based archival file for macromolecular structures. *J. Mol. Biol.*. 112, 535-542.

Burley, S.K. & Petsko, G.A. (1986) Amino-aromatic interactions in proteins. *FEBS Letters* **203**, 139-143.

D'Alessio, G., DiDonato, A., Parente, A. & Piccoli, R. (1991) Seminal RNase: a unique member of the ribonuclease superfamily. *Trends Biochem. Sci.* **16**, 104-106.

Durack, D.T., Ackerman, S.J., Loegering, D.A., & Gleich, G.J. (1981). Purification of human eosinophil-derived neurotoxin. *Proc. Natl. Acad. Sci. USA* **78**, 5165-5169.

Fontecilla-Camps, J.C., de Llorens, R., le Du, M.H. & Cuchillo, C.M. (1994) Crystal structure of ribonuclease A-d(ApTpApApGp) complex. *J. Biol. Chem.* **269**, 21526-21531.

Fredens, K., Dahl, R. & Venge, P. (1982) The Gordon phenomenon induced by eosinophil cationic protein and eosinophil protein X. *J. Allergy Clin. Immunol.* **70**, 361-366.

Gleich, G.J., Loegering, D.A., Bell, M.P., Checkell, J.L., Ackerman, S.J. & McKean, D.J. (1986) Biochemical and functional similarities between human eosinophil-derived neurotoxin and eosinophil cationic protein: Homology with ribonuclease. *Proc. Natl. Acad. Sci.* **83**, 3146-3150.

Hallahan, T.W., Shapiro, R. & Vallee, B.L. (1991) Dual site model for the organogenic activity of angiogenin. *Proc. Natl. Acad. Sci.* **88**, 2222-2226.

Hamlin, R. (1985). Multiwire area X-ray diffractometers. *Methods in Enzymol.* **114**, 416-452.

Hofsteenge, J., Muller, D.R., de Beer, T., Loffler, A., Richter, W.J. & Vliegenthart, J.F.G. (1994) New type of linkage between a carbohydrate and a protein: C-glycosylation of a specific tryptophan residue in human RNase U_S *Biochemistry* **33**, 13524-13530.

Howard, J., Nielson, C. & Xuong, Ng-H. (1985). Software for a diffractometer with multiwire area detector. *Meth. Enzym.* **114**, 452-472.

Jancarik J & Kim SH. 1991. Sparse matrix sampling: a screening method for crystallization of proteins. *J. Appl. Cryst.* **24**:409-411.

Jones, T.A. 1985. Interactive computer graphics: FRODO. *Methods Enzymol.* **115**:157-171.

Jones, T. A., Zou, J. Y., Cowan, S. W., Kjeldgaard, M. (1991) Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Cryst.* **A47**:110-119.

Kabsch, W. (1978) A discussion of the solution for the best rotation to relate two sets of vectors. *Acta Cryst.* **A34**:827-828.

Kraulis, P.J. (1991). Molscript: a program to produce both detailed and schematic plots of protein structures. *J. Appl. Cryst.* **24**, 946-950.

Laskowski, R. A., MacArthur, M. W., Moss, D. S., Thornton, J. M. (1993) PROCHECK: a program to check the stereochemical quality of protein structures. J. *Appl. Cryst.* **26**:283-291.

Lee, F.S., & Vallee, B.L. (1993) Structure and action of mammalian ribonuclease (angiogenin). *Prog. Nucl. Acid Res. Mol. Biol.* **44**, 1-30.

Lesczyznski, J.L. & Rose, G. (1986) Loops in globular proteins: A novel category of secondary structure. *Science* **234**, 849-856.

Matthews, B.W. (1968). Solvent content of protein crystals. *J. Mol. Biol.* **33**, 491-497.

Mikulski, S.M., Ardelt, W., Shogen, K., Bernstein, E.H. & Menduke, H. (1990a). Striking increase of survival of mice bearing M109 Madison carcinoma treated with a novel protein from amphibian embryos. *J. Natl. Cancer Inst.* **182**, 151-153.

Mosimann, S.C., Ardelt, W. & James, M.N.G. 1994. Refined 1.7 Å X-ray crystallographic structure of P-30 protein, an amphibian ribonuclease with anti-tumor activity. *J. Mol. Biol.* **236**:1141-1153.

Mosimann, S.C., Meleshko, R. & James, M.N.G. (1995) Assessment of the comparative molecular modeling of tertiary structures of proteins. *Proteins: Struct. Func. Genet.* (in press).

Navaza, J. AMoRe: An automated package for molecular replacement. Acta Cryst. A50:157-163, 1994.

Newton, D.L., Walbridge, S., Mikulski, S.M., Ardelt, W., Shogen, K., Ackerman, S.J., Ryback, S.M. & Youle, R.J. (1994) Toxicity of an antitumor ribonuclease to Purkinje Neurons. *J. Neuroscience* **14**, 538-544.

Newton, D.L., Nicholls, P., Ryback, S.M. & Youle, R.J. (1994) Expression and characterization of recombinant human eosinophil-derived neurotoxin and eosinophil-derived neurotoxin-anti-transferrin receptor sFv *J. Biol. Chem.* **269**, 26739-26745.

Richards, F.M. & Wykoff, H.W. (1971). Atlas of molecular structures in biology (Phillips, D.C. & Richards, F.M., eds.). vol. 1, RNase A, Clarendon Press, Oxford.

Ryback, S.M., Saxena, S.S., Ackerman, E.J. & Youle, R.J. (1991). Cytotoxic potential of ribonuclease and ribonuclease hybrid proteins. *J. Biol. Chem.* **266**, 21202-21207.

Saxena, S., Ryback, S , Winkler, G., Meade, H., McGray, P., Youle, R. & Ackerman, E. (1991). Comparison of RNases and toxins upon injection into *Xenopus Oocytes. J. Biol. Chem.* **266**, 21208-21214.

Sibanda, B. & Thornton, J.M. (1985). β-Hairpin families in globular proteins. *Nature* **316**, 170-174.

Sierakowska, H. & Shugar, D. (1977) Mammalian nucleolytic enzymes. *Prog. Nucleic Acid Res. molec. Biol.* **20**,59-130.

Titani, K., Takio, K., Kuwada, M., Nitta, K., Sakakibara, F., Kawauchi, H., Takayanagi, G, & Hakomori, S. (1987) Amino acid sequence of sialic acid binding lectin from frog (*Rana catesbeiana)* eggs. *Biochemistry* **26**,2189-2194.

Tronrud, D.E. (1992) Conjugate-direction minimization: an improved method for the refinement of macromolecules. *Acta Crsyt.* **A48**,912-916.

Wu, Y., Mikulski, S.M., Ardelt, W., Ryback, S.M. & Youle, R.J. (1993). A cytotoxic ribonucleases: Study of the mechanism of Onconase cytotoxicity. *J. Biol. Chem.* **268**, 10686-10693.

Youle, R.J., Newton, D., Wu, Y., Gadino, M. & Ryback, S.M. (1993) Cytotoxic ribonucleases and chimeras in cancer therapy. *Critical Rev. Therapeutic Drug Carrier Systems* **10**, 1-28.

Zegers, I., Maes, D., Thi, M. H. D., Wyns, L., Poortmans, F., Palmer, R. (1994) The structure of RNase A complexed with 3'-CMP and d(CpA): Active site conformation and conserved water molecules. Protein Science **3**:2322-2339.

CHAPTER 6

DISCUSSION

ONC and EDN are cytotoxic members of the pyrimidine specific RNase superfamily of vertebrates. They have therapeutic potential and play a role in various human pathologies, respectively. To further our understanding of the biological properties of these enzymes, the molecular structures of ONC, rONC and EDN were determined and refined at high resolution using X-ray crystallographic techniques. They represent the first reported structures of amphibian and non-secretory pyrimidine-specific RNases. This is significant as the homologous RNase A and angiogenin enzymes are not cytotoxins.

The overall folds of ONC and rEDN are related to those observed in crystal structures of RNase A (Wlodawer et al, 1983; Zegers *et al.*, 1994) and angiogenin (Acharya et al., 1994). These enzymes are composed of an N-terminal α-helix and two similarly folded lobes. Lobe 1 contains strands β1, β3 and β4 and helix α2 while lobe 2 contains strands β2, β5 and β6 and helix α3. Each lobe forms a three-stranded antiparallel β-sheet and the α-helix packs against the β-sheet. In addition, the α-helix of each lobe contributes a conserved disulfide bridge to the central strand of their respective β-sheets. The similarity in the structure of these two lobes suggests a gene duplication event gave rise to the original RNase.

While these enzymes have the same overall fold, their detailed molecular structures are significantly different. In particular, the N-terminus, the loop structures following the three α-helices and the omega loops following the central strands of the two β-sheets have very different conformations in each of these enzymes. In addition, there are a number of non-conservative amino acid replacements among structurally equivalent residues. These differences are expected to be responsible for the biological properties of these enzymes.

At present, our lack of knowledge regarding the identity and nature of the receptors of ONC and EDN prevent us from identifying the residues involved in the cell binding activity of these enzymes. Future studies using site directed mutants of rONC and rEDN will be aimed at identifying residues that facilitate the cytotoxic mechanism of these enzymes.

B_1 Subsite of Pyrimidine Specific RNases

The active site of the pyrimidine specific RNases is located at the junction of the N-terminal α-helix, lobe 1 and lobe 2. Together, they form a cleft that runs parallel to strand β6. At the base of the cleft are two invariant residues; a histidine and a threonine. In the B_1 subsite the N and $O^{\gamma 1}$ atoms of the invariant Thr complement the hydrogen bonding potential of the O(2) and N(3) atoms of the pyrimidine base, respectively. As a consequence, the differences in the pyrimidine-base specificity of these two enzymes arise from differences in the binding of the N(4) (cytosine)/O(4) (uracil) atom. In the RNase A-inhibitor structures, direct hydrogen bonds between N(4) and Ser123$O^{\gamma 1}$ (Richards & Wykoff, 1971) and water-mediated interactions between O(4) and Ser123$O^{\gamma 1}$ (Wlodawer *et al.*, 1983; Borah *et al.*, 1985) have been reported. Since Ser123$O^{\gamma 1}$ and water molecules can either accept or donate hydrogen bonds, they account for the pyrimidine base preference of RNase A. The residues of ONC, EDN and angiogenin that are equivalent Ser123 of RNase A are Val101, Ile133 and Ser124, respectively. The ONC-d(UpG) model (Chapter 4) and the rEDN-$d(Tp)_3$ model (Chapter 5) suggest Lys33 of ONC and Gln40 of EDN replace the Ser123 interaction with the B_1 base observed in RNase A complex structures. Lys33 of ONC and Gln40 of EDN are structurally equivalent residues that account for the observed specificity of their respective enzymes. In RNase A, the counterpart to these residues is Val43. As a result, two non-conservative substitutions are required to change the specificity of ONC to that of RNase A or vice versa. It is interesting that angiogenin has an Arg that is equivalent to Lys33 of ONC and retains a Ser that is equivalent to Ser123 of RNase A. Both angiogenin residues can participate in hydrogen bonding interactions with

the O(4) (uracil)/N(4) (cytosine) atoms of the pyrimidine base and angiogenin has a specificity that is intermediate to that observed for ONC and RNase A (Table 1.4).

B_2 Subsite of Pyrimidine Specific RNases

The B_2 subsite of all pyrimidine specific RNases includes a Glu (RNase A; Glu111) or an Asp (EDN; Asp112) that are structurally equivalent. In RNase A, EDN and angiogenin there are several residues from the loop l4 that can participate in substrate binding. In ONC a large portion of this loop has been deleted (Chapter 2, 3). As a consequence, the loop l4 and the B_2 subsite of ONC have very different conformation when compared to other pyrimidine specific RNases. In angiogenin, residues of the loop l4 contribute to its angiogenic potential. The loop l4 of angiogenin has been replaced with the equivalent loop of RNase A and resulting enzyme shows increased enzymatic activity and a greater RNase A-like character. Moreover, the chimeric enzyme loses its angiogenic activity (Hallahan *et al.*, 1991). As a result, it has been proposed that the loop l4 is a module that controls the enzymatic activity and biological function of pyrimidine specific RNases. If this is correct, the deletion of this loop may confer the guanine specificity of ONC to other members of the superfamily.

rONC

In the X-ray crystallographic structure of ONC (Chapter 3), the N-terminal pyroglutamyl residue folds back against the N-terminal α-helix and forms a hydrogen bond with Lys9 at the enzymes active site. An equivalent pair of residues is found in each of the amphibian RNases and in the iguana RNase (Zhao *et al.*, 1994). In all mammalian members of this superfamily the residue equivalent to Lys9 is a Gln. This suggests the Pyr1 interaction with Lys9 is an important feature of the non-mammalian enzymes. The discovery that rONC with an N-terminal sequence $^{f}Met_{-1}Glu_{1}Asp_{2}\cdots$ lacks significant RNase activity and cytotoxicity confirms Pyr1 has a role in the RNase activity of the

enzyme (Wu, Y. & Youle, R.J., unpublished results). At present several approaches to restoring the activity of rONC are being explored. These include; the generation of Pyr1 by chemical means after CNBr or enzymatic removal of $^{f}Met_{-1}$ and the construction of point mutants that would replace the function of Pyr1 in the R_2 subsite (Chapter 4). In the structure of ONC, Pyr1 packs against Thr5 of the N-terminal α-helix. In all mammalian pyrimidine specific RNases the residues equivalent to Thr5 of ONC are lysine (RNase A), tryptophan (EDN) or histidine (angiogenin). The relatively large side chains of these amino acids sterically clash with Pyr1 when modeled in place of Thr5. As a result we have suggested a mutation of Thr5 to Trp to replace Pyr1 in the R_2 subsite and an additional mutation, Asp2 to Trp. By changing Asp2, the negative charge of the β carboxyl is removed from the R_2 subsite and potentially unfavorable electrostatic interactions with adjacent phosphate subsites are avoided (Chapter 4).

P_{-1} Subsite of rEDN

The structure also contains two discreet sulfate anion binding sites. The active site anion is non-covalently bound in a similar position and conformation in each of these structures. The SO_4(B) anion occupies a site that may correspond to the P_{-1} subsite. SO_4(B) hydrogen bonds to the main-chain and side-chain atoms of Asn39. In RNase A, the residue equivalent to Asn39 is Pro42. Clearly, Pro42 can make neither of the Asn39 interactions observed in the rEDN structure. In a tetra deoxythymidylic acid-RNase A complex (Birdsall et al., 1992) the P_{-1} phosphate makes van der Waal's contacts with Pro42. In the absence of Pro42, RNase A may bind the P_{-1} phosphate as observed in rEDN. Since the position equivalent to Pro42 (RNase A) is not highly conserved within the superfamily, the SO_4(B) binding in rEDN may be more representative of how other RNase bind extended substrates.

Comparative Molecular Modeling

The comparative molecular modeling study of ONC (Chapter 2) demonstrates ONC can adopt a fold related to that observed in RNase. The model correctly predicted the conformation of residue Ser54-Leu58 justifying a revision in the existing automated sequence alignment. The original sequence alignment predicted the deletion of strand β2 and aligned these residues with a large loop structure (Ardelt *et al.*, 1991). The comparative molecular modeling did not correctly predict the structures of loops with large insertions and deletions and failed to predict the conformation of the N-terminus and helix α2. As evidence of the differences in the model and observed structures of ONC, the model was unable to provide a molecular replacement solution. The correct rotation function solution does not appear among the top 100 rotation function peaks when using the model. Previous and subsequent comparative molecular model assessments have suggested these loop structures are exceedingly difficult to predict when they are the sites of insertions and deletions and have little or no amino acid sequence identity (Greer, 1985; Mosimann *et al.*, 1995). The helix α2 is flanked loops with large deletions and is more than 2Å from its predicted position. The modest amino acid sequence identity shared by these enzymes is restricted to disulfide bridges that help define the tertiary structure of these enzymes, residues known to be important for catalysis and several residues of the hydrophobic core (Table 1.2). As a result, the loop structures of RNase A are a poor model for the observed conformation of the ONC and rEDN loops.

REFERENCES

Acharya, K.R., Shapiro, R., Allen, S.C., Riordan, J.F., & Vallee, B.L. (1994). Crystal structure of human angiogenin reveals the structural basis for its functional divergence from ribonuclease. *Proc. Natl. Acad. Sci. U.S.A.* **91**:2915-2919.

Ardelt, W., Mikulski, S.M., & Shogen, K. 1991. Amino acid sequence of an anti-tumor protein from *Rana Pipiens* oocytes and early embryos. *J. Biol. Chem. 266*:245-251.

Birdsall, D.L. & McPherson, A. (1992). Crystal structure disposition of thymidylic acid tetramer in complex with ribonuclease A. *J. Biol. Chem.* **267**, 22230-22236.

Borah, B., Chen, C., Egan, W., Miller, M., Wlodawer, A., & Cohen, J.S. 1985. Nuclear magnetic resonance and neutron diffraction studies of the complex of ribonuclease A with uridine vanadate, a transition state analogue. *Biochemistry 24*:2058-2067.

Greer, J. (1985). Protein structure and function by comparative model building. *Ann. N.Y. Acad. Sci.* **439**, 44-63.

Hallahan, T.W., Shapiro, R. & Vallee, B.L. (1991) Dual site model for the organogenic activity of angiogenin. *Proc. Natl. Acad. Sci.* **88**, 2222-2226.

Mosimann, S.C., Meleshko, R. & James, M.N.G. (1995) Assessment of the comparative molecular modeling of tertiary structures of proteins. *Proteins: Struct. Func. Genet.* (in press).

Richards, F.M., & Wyckoff, H.W. 1971. The Enzymes (Boyer P. ed.) vol. 4:647-806, Academic Press, New York.

Titani, K., Takio, K., Kuwada, M., Nitta, K., Sakakibara, F., Kawauchi, H., Takayanagi, G. &, Hakomori, S. (1987) Amino acid sequence of sialic acid binding lectin from frog (*Rana catesbeiana)* eggs. *Biochemistry* **26**,2189-2194.

Zegers, I., Maes, D., Thi, M. H. D., Wyns, L., Poortmans, F., Palmer, R. (1994) The

structure of RNase A complexed with 3'-CMP and d(CpA): Active site conformation and conserved water molecules. Protein Science 3:2322-2339.

Zhao, W., Beintema, J.J. & Hofsteenge, J. (1994) The amino acid sequence of iguana pancreatic ribonuclease. *Eur. J. Biochem.* **219**, 641-646.